GRAMÁTICA PEDAGÓGICA DO PORTUGUÊS BRASILEIRO

Marcos Bagno

parábola

201437

Marcos Bagno

UNIVERSIDADE FEDERAL
Biblioteca Comunitária
S. CAR

151961

GRAMÁTICA PEDAGÓGICA DO PORTUGUÊS BRASILEIRO

parábola

DOAÇÃO R-T
DE Jasira R Rosa
EM 28/6/2019
VALOR EST. R$ 176,88

Clas. G469.5
Cutt. B134g
2. ed.
Tombo 15.1961

Projeto gráfico: Andreia Custodio
Foto da Capa: © Arquivo Público do Distrito Federal
Ilustrações: Miguel Bezerra
Editor: Marcos Marcionilo
Conselho Editorial: Ana Stahl Zilles (Unisinos)
Angela Paiva Dionisio (UFPE)
Carlos Alberto Faraco (UFPR)
Egon de Oliveira Rangel (PUC-SP)
Henrique Monteagudo (Universidade de Santiago de Compostela)
José Ribamar Lopes Batista Jr. (UFPI/CTF/LPT)
Kanavillil Rajagopalan (UNICAMP)
Marcos Bagno (UnB)
Maria Marta Pereira Scherre (UFES)
Rachel Gazolla de Andrade (PUC-SP)
Roberto Mulinacci (Universidade de Bolonha)
Roxane Rojo (UNICAMP)
Salma Tannus Muchail (PUC-SP)
Sirio Possenti (UNICAMP)
Stella Maris Bortoni-Ricardo (UnB)

CIP-BRASIL. CATALOGAÇÃO NA FONTE
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ

B134g

Bagno, Marcos, 1961-
Gramática pedagógica do português brasileiro / Marcos Bagno.
- São Paulo: Parábola Editorial, 2012.

1056p. ; 24 cm. (Referenda ; 1)
ISBN 978-85-7934-037-6

1. Língua portuguesa - Gramática - Estudo e ensino. I. Título. II. Série.

11-6457. CDD: 469.5
CDU: 811.134.3'36

Direitos reservados à
Parábola Editorial
Rua Dr. Mário Vicente, 394 – Ipiranga
04270-000 São Paulo, SP
pabx: [11] 5061-9262 | 5061-8075 | fax: [11] 2589-9263
home page: www.parabolaeditorial.com.br
e-mail: parabola@parabolaeditorial.com.br

Todos os direitos reservados. Nenhuma parte desta obra pode ser reproduzida ou transmitida por qualquer forma e/ou quaisquer meios (eletrônico ou mecânico, incluindo fotocópia e gravação) ou arquivada em qualquer sistema ou banco de dados sem permissão por escrito da Parábola Editorial Ltda.

ISBN: 978-85-7934-037-6

1ª edição, 2ª reimpressão: junho de 2016

© do texto: Marcos Bagno, 2012
© da edição: PARÁBOLA EDITORIAL, São Paulo, janeiro de 2012.

"O discurso herético deve contribuir não somente para romper com a adesão ao mundo do senso comum, professando publicamente a ruptura com a ordem ordinária, mas também produzir um novo senso comum e nele introduzir as práticas e as experiências até então tácitas ou recalcadas de todo um grupo, agora investidas da legitimidade conferida pela manifestação pública e pelo reconhecimento coletivo."

PIERRE BOURDIEU (1996: 119)

para Orlene Lúcia de Sabóia Carvalho,
amizade maior, maior que a amizade

sumário

Agradecimentos, 11
Aviso aos navegantes, 13
Abreviaturas e símbolos, 15
Símbolos fonéticos, 16
Introdução: gramática, a quem será que se destina?, 19

LIVRO I - EPISTEMOLOGIA DO PORTUGUÊS BRASILEIRO

1. A CAVERNA IMPLODIDA - por uma concepção não-platônica de língua, 37
2. O DEVANEIO DA LÍNGUA PRIMITIVA - colonialismo, racismo e preconceito linguístico, 81
3. DE LÍNGUA MATERNA A LÍNGUA PATERNA - do vernáculo à normatização, 99

LIVRO II - HISTÓRIA DO PORTUGUÊS BRASILEIRO

4. NADA SERÁ COMO ANTES - a mudança linguística, 115
5. DO GALEGO AO BRASILEIRO - história da nossa língua, 201
6. RAIZES DESTERRADAS - formação do léxico portugalego, 255

LIVRO IIII - MULTIMÍDIA DO PORTUGUÊS BRASILEIRO

7. OS SONS E OS SÉCULOS - fonologia da nossa língua, 291
8. RUIDOS E RABISCOS - língua falada e língua escrita, 343

LIVRO IV - LEXICOGRAMÁTICA DO PORTUGUÊS BRASILEIRO

9. UM PRESENTE DE GREGO - história das classes gramaticais, 401
10. UNIVERSAIS E BRASILEIROS - conceitos importantes para entender a gramática, 431
11. AS PALAVRAS, AS COISAS E AS NÃO COISAS - características lexicogramaticais do português brasileiro, 495
12. NO PRINCÍPIO ERA... - o verbo, 507
13. UMA ROSA É UMA ROSA É UMA ROSA - os nomes, 663
14. ENTRE DOIS AMORES - os verbinominais, 715
15. QUESTÕES PESSOAIS - os índices de pessoa, 737
16. DE MONSTROS E DEMONSTRAÇÕES - os mostrativos, 773
17. TODOS, ALGUNS, NENHUM - os quantificadores, 825
18. SEMPRE CABE MAIS UM - os adverbios, 831
19. PEQUENAS NOTÁVEIS - as preposições , 853
20. OS NÓS E OS NEXOS - as conjunções e companhia ilimitada, 881

LIVRO V - DIDÁTICA DO PORTUGUÊS BRASILEIRO

21. ERREI, SIM - a hipercorreção e suas consequências, 933
22. O QUE (NÃO) ENSINAR NA ESCOLA - por uma educação linguística realista, 983

Bibliografia, 1011
Índice de assuntos, 1024
Índice de nomes, 1037
Índice geral, 1042

agradecimentos

Em primeiríssimo lugar, ao meu querido amigo e editor MARCOS MARCIONILO, que se apaixonou pelo projeto e ficou tão ansioso quanto eu para vê-lo concluído. E como tudo o que ele faz tem o apoio imprescindível de ANDRÉIA CUSTÓDIO, minha gratidão se estende a ela também.

Esse livro é profundamente devedor a ATALIBA TEIXEIRA DE CASTILHO, mestre generoso, que há tantas décadas vem escavando o português brasileiro para nos presentear com seu ouro e seus cristais — foi a leitura de sua monumental gramática que me instigou a produzir essa aqui, que deve muito à troca de mensagens que forcei ele a manter comigo em plena época de festas de fim de ano e ano novo adentro. A ele também devo a cessão do *corpus* compartilhado do projeto NURC, material indispensável para qualquer pesquisa sobre a verdadeira língua dos brasileiros letrados.

Muito tenho aprendido com MARIA EUGÊNIA LAMOGLIA DUARTE, minha cicerone pelas belezas do Rio de Janeiro e pelas trilhas emocionantes da sintaxe. Nada se compara a uma bela aula sobre verbos numa varanda que se debruça sobre a deslumbrante paisagem da Lagoa e da Mata Atlântica.

Agradeço à minha amiga maior, ORLENE CARVALHO, pela capacidade de manter um segredo e aparecer em momentos estratégicos para me salvar de mim mesmo. Se não me engano, o nome disso é amor.

Das margens do Rio da Prata, meu querido FEDERICO POLASTRI me incentivou a escrever esse manual. Espero que também possa servir para os que amam a nossa língua e querem aprendê-la e ensiná-la aos que se interessam por ela. *Gracias*, Fede.

Meu filho MIGUEL BEZERRA traduziu em ilustrações as ideias que minhas mãos sempre foram inábeis para transformar em imagens convincentes. Obrigado, Mig.

Sou muito grato ao Arquivo Público do Distrito Federal — na pessoa do Sr. MARCELO DURÃES — e ao Centro de Documentação da Universidade de Brasília — na pessoa da Sra. LÚCIA HOCHREITER — que nos cederam muito gentilmente as fotografias que ilustram esse livro.

Por fim, agradeço a DENNYS DA SILVA REIS, por ser quem é e ter estado ao meu lado durante todo o tempo e o espaço exigidos para uma tarefa desse tamanho. Sem ele teria sido difícil compilar sozinho toda a bibliografia que serviu de base a esse trabalho.

Ερως, αγάπη, φίλος.

Brasília, agosto de 2011.

aviso aos navegantes

Para uma leitura não ingênua desse livro, é importante deixar claros alguns de seus pressupostos, de modo a evitar futuros mal-entendidos e eventuais cobranças do que não foi prometido. Sendo assim, essa obra

1. é uma GRAMÁTICA, na medida em que pretende examinar e descrever o funcionamento de uma língua específica, o português brasileiro contemporâneo. Esse exame-descrição, no entanto, não é exaustivo,

pois o mais importante nesse projeto é destacar as especificidades da nossa língua, as que tornam ela diferente das outras línguas de seu grupo (o portugalego, ver capítulo 4) e também das demais línguas da família românica;

2. é PROPOSITIVA, porque não se limita a descrever ou a expor o português brasileiro, mas propõe efetivamente a plena aceitação de novas regras gramaticais que já pertencem à nossa língua há muito tempo e, por isso, devem fazer parte do ensino sistemático da língua. Ela formula um "discurso herético", no sentido conferido à expressão por Pierre Bourdieu no trecho que lhe serve de epígrafe;

3. é PEDAGÓGICA, porque foi pensada para colaborar com a formação docente que, no Brasil, é reconhecidamente falha e precária. Nossos cursos de Letras (a começar pelo nome) se vinculam a um ideário cultural obsoleto, enraizado na sociedade burguesa do século XIX. Por isso, eles deixam de oferecer aos estudantes uma série de conhecimentos fundamentais enquanto, por outro lado, desperdiçam tempo com a transmissão de conteúdos irrelevantes para quem vai exercer a profissão docente. Basta perguntar a professoras e professores na ativa ou em formação se sabem, por exemplo, o que é *gramaticalização* ou se ao menos já ouviram falar disso;

4. é um projeto EPISTEMOLÓGICO porque traz explícita uma teoria do conhecimento, destinada a fundamentar os posicionamentos francamente assumidos ao longo de todo o texto (ver capítulo 1);

5. é POLÍTICO-IDEOLÓGICA porque é um produto humano e não existe produto humano que não se configure, consciente ou inconscientemente, como uma tomada de posição política inspirada por uma ou mais ideologias; o mito da ciência "neutra" não tem mais lugar na era em que vivemos. Assim, essa obra milita a favor do reconhecimento do português brasileiro como uma língua plena, autônoma, que deve se orientar por seus próprios princípios de funcionamento e não por uma tradição gramatical voltada exclusivamente para o português europeu literário antigo. Essa militância se traduz no emprego consciente de formas linguísticas há muito tempo incorporadas à gramática do português brasileiro, mas que ainda são alvo da perseguição dos puristas mais empedernidos. Por isso, ninguém se assuste ao topar com construções do tipo *"nos grupos que fazemos parte"*, ou *"tem muitos problemas nessa descrição"*, ou *"tendo transformado ela numa regra"*, ou *"não se conhece as origens exatas dessas palavras"*, entre outras;

6. é TEÓRICA na medida em que discute, refuta ou abraça propostas anteriores de descrição da língua e em que propõe novas análises, definições e conceitos;

7. é HISTÓRICA porque rejeita a tradicional separação entre diacronia e sincronia e assume o fenômeno linguístico como eminentemente pancrônico, variável e mutante. Desse modo, o recurso às transformações ocorridas na(s) língua(s) ao longo do tempo é indispensável para o (re)conhecimento preciso do que ocorre aqui e agora.

Com isso em mente, espero que os eventuais leitores dessa obra tirem algum proveito de um trabalho sincero, árduo, mas também muito prazeroso.

Fortuna audaces iuvat.

abreviaturas e símbolos

[▸ 123] = remete à página indicada
A < B = a provém de b
A > B = a gerou b
a.C. – antes de Cristo
AAVE – *Afro-American Vernacular English*
AUM = aumentativo
BEV – *Black English Vernacular*
CAUS = causativo
CON = consoante
COMP OBL = complemento oblíquo
cp. – compare-se
d.C. – depois de Cristo
DIM = diminutivo
+GER = + gerúndio
GTM – gêneros textuais mais monitorados
GU = gramática universal
IMPERF = imperfeito
INDIC = indicativo
+INF = + infinitivo
IP – índices de pessoa
NNE = norte-nordeste
ÑP – não-pessoa
NURC – Norma Urbana Culta (Projeto)
OB = objeto
OBJ D = objeto direto
OBJ IND = objeto indireto
PB – português brasileiro
PL = plural
PRET = pretérito
QI – quantificadores indefinidos
S = sentença
SENS = sensitivo
SING = singular
SN = sintagma nominal
SSE = sul-sudeste
SUJ = sujeito
SV = sintagma verbal
TGP – tradição gramatical do português
V = verbo
VGB – vernáculo geral brasileiro
VLE – variedades linguísticas estigmatizadas
VOG = vogal
VUP – variedades urbanas de prestígio
° – antes de um enunciado, indica sua agramaticalidade (inaceitabilidade)
° – na linguística histórica, representa forma hipotética reconstruída

símbolos fonéticos

Os exemplos, salvo indicação em contrário, se baseiam na pronúncia "neutra" descrita em [▶339]

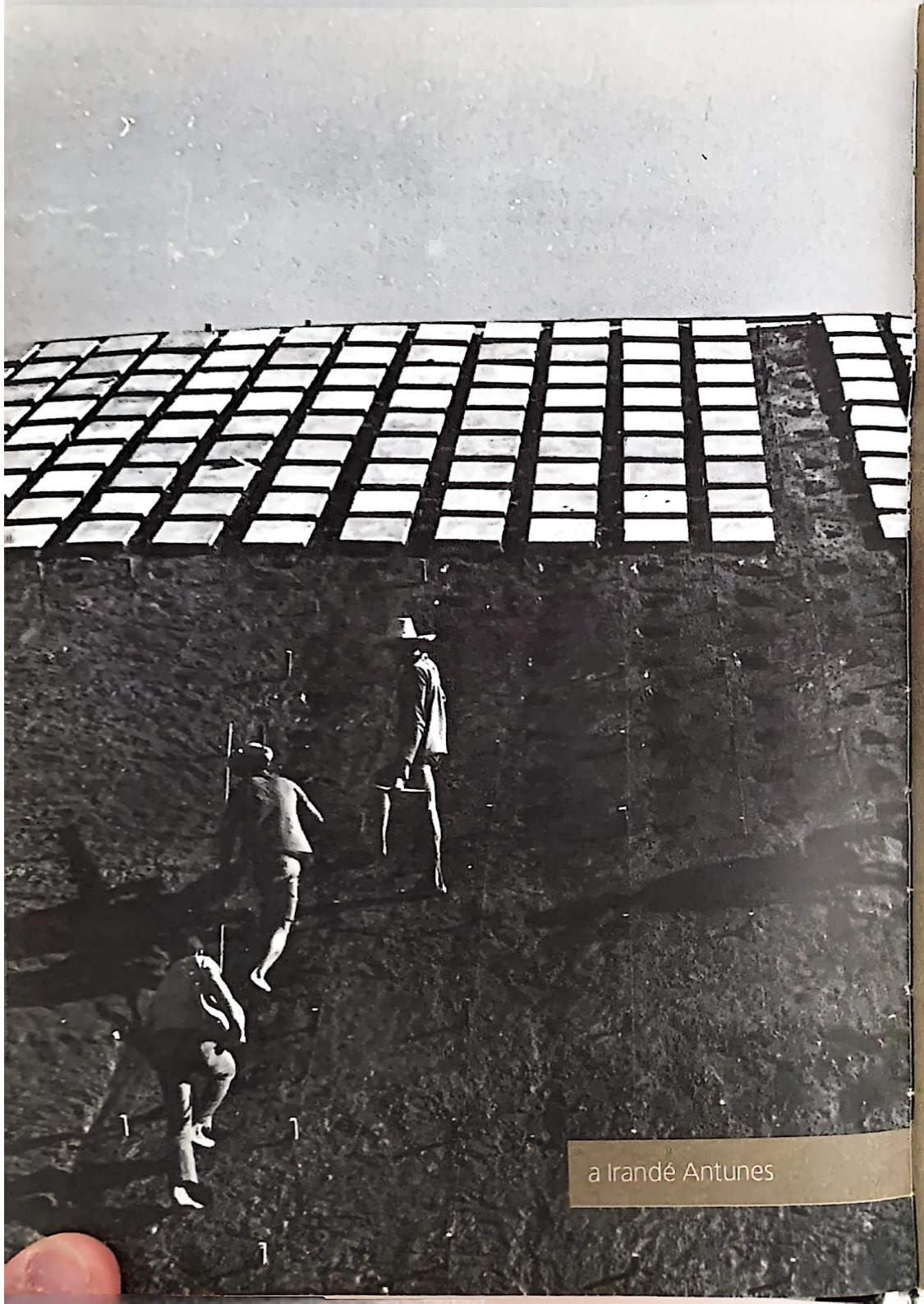
[ˈ] precede a sílaba tônica como em [ˈfala]

1. VOGAIS E SEMIVOGAIS

[a] como em **pá**
[ɐ] como no port. eur. **madeira**
[e] como em **bebê**
[ə] como no port. eur. **noite**; franc. **devoir**; ingl. **first**
[ɛ] como em **pé**
[i] como em **vi**
[o] como em **vovô**
[ɔ] como em **pó**
[u] como em **urubu**
[y] como em francês **tu**
[œ] como em francês **fleur**
[ø] como em francês **deux**
[ã] como em **manhã**
[ẽ] como em **vento**
[ĩ] como em **fim**
[õ] como em **som**
[ũ] como em **mundo**
[ɪ] como em **pai**
[ʊ] como em **pau**; **mau**; **mal**

2. CONSOANTES

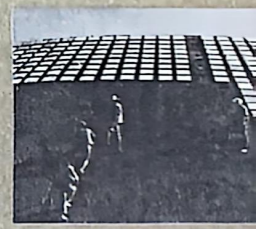
[b] como em **baba**
[β] como em port. eur., galego ou esp. **acabar, abrir** [▶137]
[k] como em **cuca**
[kʷ] como em **quadro**
[d] como em **dado**
[ð] como em port. eur., galego e esp. **acabado, padre** [▶137]
[ʤ] como em **dia** (pronúncia carioca)
[f] como em **fofo**
[g] como em **gago**
[ɣ] como em port. eur., galego e esp. **lago, alegre** [▶137]
[gʷ] como em **água**
[ʒ] como em **já**
[l] como em **lá**
[ɬ] como no port. eur. **mel**, no ing. **well**
[ʎ] como em **palha**
[m] como em **mama**
[n] como em **nó**
[ɲ] como em **manha**
[p] como em **pó**
[ɾ] como em **puro**
[r̃] como em italiano e espanhol **carro** [▶323]
[x] pronúncia característica, por exemplo, da variedade carioca em **rio** e **terra**. Ocorre fricção audível na região velar. Se assemelha ao que se escreve com *j* em castelhano (*Juan, caja, jamón*). [▶323]
[ɣ] correspondente vozeada da consoante anterior e ocorre em final de sílaba seguida de consoante vozeada como em **perda**. [▶323]
[ɹ] "R caipira" em **porta** [▶323].
[h] pronúncia característica, por exemplo, da variedade mineira de Belo Horizonte em **rio** e **terra**. Não ocorre fricção audível na região velar. [▶323]
[ɦ] correspondente vozeada da anterior, ocorre em final de sílaba seguida de consoante vozeada, como em **perda**. [▶323]
[s] como em **só**
[ʃ] como em **xixi**
[t] como em **tatu**
[ʧ] como em **tcheco**
[θ] como em inglês **think**
[v] como em **vovô**
[z] como em **doze**



a Irandé Antunes

introdução

gramática: a quem será que se destina?



Faz um bom tempo já que se firmou entre os pesquisadores da área da educação linguística a convicção de que a função primordial da escola, no que diz respeito à pedagogia de língua materna, é promover o letramento de seus aprendizes. E para essa promoção do letramento, as atividades fundamentais são a *leitura* e a *escrita*, com foco na diversidade de *gêneros textuais* que circulam na sociedade.

Além da leitura e da escrita, também tem espaço em sala de aula para a *reflexão sobre a língua e a linguagem*. Essa reflexão deve ser feita primordialmente através das chamadas *atividades epilinguísticas*, aquelas que não recorrem à nomenclatura

técnica (a *metalinguagem*), de modo a permitir o percurso uso→reflexão→uso. Isso, logo de saída, implica que tais atividades só podem ser feitas a partir de textos autênticos, falados e escritos, dos quais se possa depreender o funcionamento da língua na construção dos sentidos. O enfoque deve ser, portanto, essencialmente semântico-pragmático-discursivo: as reflexões sobre os aspectos especificamente *gramaticais* precisam ser lançadas contra esse pano de fundo semântico-pragmático-discursivo, de modo a conscientizar o aprendiz de que os recursos disponíveis na língua são ativados essencialmente para *a produção de sentido e a interação social*.

É do uso que se depreende a gramática, é do discurso que se chega nas regularidades (sempre instáveis e provisórias) da língua — uma distinção, é claro, que tem aqui uma perspectiva apenas pedagógica, já que na prática social mais ampla discurso e sistema (ou uso e gramática) interagem sem cessar, são indissociáveis, tanto quanto o oxigênio e o hidrogênio da água: são os usos frequentes e regulares de determinada forma linguística que acabam por transformá-la em regra gramatical, assim como são as regras gramaticais as condicionadoras dos usos linguísticos. Dado que só existe língua se existirem falantes dessa língua, ou seja, só existe língua em uso, a prática da linguagem como atividade constitutiva da própria natureza humana (natureza cognitiva e sociocultural) é que ditará os rumos da gramática da língua, num processo cíclico e permanente, que só se interrompe quando e se deixarem de existir falantes da língua.

Por isso, partindo da convicção de que não se deve fazer um ensino explícito, técnico e taxonômico de gramática na educação básica, esse livro tem, primordialmente, no seu horizonte de leitores potenciais, as professoras e os professores em formação ou já formados que exercem o magistério no ensino fundamental e/ou médio e na educação de jovens ou adultos, ou que se preparam para essa tarefa. Isso significa que partimos do pressuposto de que essas pessoas

I. já têm um conhecimento básico da linguística moderna, suas principais correntes teóricas, conceitos e postulados;
II. têm familiaridade com a doutrina gramatical tradicional, seja pelo fato de terem estudado nela, seja por se guiarem por ela em sua prática pedagógica, seja por terem sido apresentadas às críticas e reformulações a que essa doutrina tem sido submetida pela linguística científica no último século e meio.

Por isso, essa gramática não é uma descrição exaustiva e detalhada do PB, mas uma exposição daquilo que constitui conhecimentos necessários para um trabalho relevante e construtivo de educação linguística. Sendo a primeira gramática propositiva de uma pedagogia do português brasileiro — no sentido de se dirigir especificamente à prática docente —, nela vou me concentrar nos aspectos mais relevantes para que professoras e professores se conscientizem dos principais traços característicos do PB, conscientização indispensável para quem se ocupa da educação linguística hoje no Brasil. Essa postura acarreta algumas decisões eminentemente políticas:

- considerar o português brasileiro como uma língua plena e autônoma (e não como uma "variedade" do português europeu), dentro de um grupo de línguas que vou chamar aqui de *portugalego* [▶202];
- assumir como válido, aceitável e correto todo e qualquer uso linguístico que já esteja plenamente incorporado ao vernáculo geral brasileiro [▶108], falado e escrito, conforme uma vasta exemplificação da língua viva que nos esforçamos aqui em apresentar;
- assumir, graças ao conhecimento desse vernáculo geral, a existência de uma norma urbana culta real, radicalmente distinta da *norma-padrão clássica*, ideal, prescritiva e totalmente desvinculada dos usos autênticos do PB;
- postular que o ensino de língua se faça com base nessa norma urbana culta real, de modo a facilitar sua aquisição por parte dos aprendizes provindos das camadas sociais usuárias de outras variedades sociolinguísticas; embora exista uma distância entre essas variedades e a norma urbana culta real, ela é muito menor do que a que existe entre essas variedades e a norma-padrão clássica, na qual nem mesmo os cidadãos urbanos mais letrados se reconhecem.

Por fim, mas não menos importante, não se deve esquecer que a população brasileira apresenta níveis baixíssimos de letramento. As pesquisas realizadas para o estabelecimento do Inaf (Índice Nacional de Alfabetismo Funcional) revelam, de modo estarrecedor, que 75% da população entre 14 e 64 anos de idade é incapaz de ler e compreender um texto de dificuldade mediana.

INAF / BRASIL - Evolução do Indicador de Alfabetismo (população de 15 a 64 anos)						
	2001 2002	2002 2003	2003 2004	2004 2005	2007	2009
Analfabeto	12%	13%	12%	11%	9%	7%
Rudimentar	27%	26%	26%	26%	25%	21%
Básico	34%	36%	37%	38%	38%	47%
Pleno	26%	25%	25%	26%	28%	25%

Essas mesmas pesquisas chegaram aos seguintes resultados:

- Entre os brasileiros que nunca foram à escola ou não chegaram a completar a 1ª série, 66% são analfabetos absolutos e 95% analfabetos funcionais.
- A maioria (54%) dos brasileiros entre 15 e 64 anos que estudaram até a 4ª série atinge no máximo o grau rudimentar de alfabetismo, ou seja, possuem no máximo a habilidade de localizar informações explícitas em textos curtos ou efetuar operações matemáticas simples, mas não são capazes de compreender textos mais longos, de localizar informações que exijam alguma inferência ou mesmo de definir uma estratégia de cálculo para a resolução de problemas.
- E ainda mais grave: 10% destes indivíduos podem ser considerados analfabetos absolutos em termos de habilidades de leitura/escrita, não conseguindo nem mesmo decodificar palavras e frases, ainda que em textos simples, ou apresentam grandes dificuldades em lidar com números em situações do cotidiano, apesar de terem cursado um a quatro anos do ensino fundamental.
- Dentre os que cursam ou cursaram da 5ª a 8ª série, apenas 15% podem ser considerados plenamente alfabetizados. Chama mais a atenção o fato de que 24% dos que completaram entre cinco e oito séries do ensino fundamental ainda permaneçam no nível rudimentar, com graves limitações tanto em termos de suas habilidades de leitura/escrita quanto em matemática.
- Somente 38% dos que cursaram alguma série ou completaram o ensino médio atingem o nível pleno de alfabetismo (esperado para 100% deste grupo).
- Somente entre os que chegaram ao ensino superior é que prevalecem (68%) os indivíduos com pleno domínio das habilidades de leitura/escrita e das habilidades matemáticas[1].

Diante desse baixíssimo grau de letramento, não fica claro e evidente que a tarefa primeira, primordial, principal, precípua — para não dizer exclusiva — da nossa escola fundamental é promover a educação linguística ininterrupta dos aprendizes para que atinjam o nível pleno de alfabetismo?

Ou será que alguém acredita que é possível levar uma pessoa a dominar plenamente as habilidades de leitura e escrita obrigando ela a decorar a suposta diferença entre *adjunto adnominal* e *complemento nominal*? Ou a fazer uma criança na 5ª série/6º ano (lembrando que, segundo as pesquisas, a maioria dos alunos nessa etapa do ensino são praticamente analfabetos) a reconhecer e rotular uma *oração subordinada substantiva objetiva direta reduzida de infinitivo*? Ou a querer

[1] Todas as informações acerca do Inaf estão disponíveis no excelente portal do Instituto Paulo Montenegro na internet: <www.ipm.org.br>. As citações que faço são extraídas do documento *Inaf Brasil 2009 — Indicador Nacional de Alfabetismo Funcional. Principais resultados*, que pode ser baixado livremente desse portal.

que essa mesma criança aprenda o que é um *fonema* e, para piorar, segundo a definição totalmente errada, do ponto de vista da linguística científica, de fonema como "som da língua"?

O escasso e precioso tempo que se passa na escola não pode ser desperdiçado com tanta coisa inútil, irrelevante e, como se não bastasse, repleta de inconsistências teóricas, de erros puros e simples, de absurdos metodológicos.

Um roteiro de estudos

Mas e se o pressuposto expresso na p. 20 não corresponder à realidade? O que fazer?

Muitos estudantes de Letras se queixam de não receber uma sólida formação de doutrina gramatical tradicional em seus cursos. E têm toda razão. Os linguistas filiados a uma perspectiva investigativa e científica frequentemente adotam um discurso pejorativo com relação à gramática tradicional, discurso que muitas vezes beira o preconceito. E não consideram pertinente abordar essa tradição em seus cursos na universidade supondo, erroneamente, que os estudantes já tiveram suficiente contato com ela durante a escolarização básica. No entanto, como patrimônio cultural do Ocidente, a gramática tradicional tem de ser muito bem conhecida por aqueles que, profissionalmente, serão confrontados a ela — cobrados para que a ensinem, desafiados a dizer por que não a ensinam, acusados de não reconhecer a suposta (e nunca comprovada) necessidade de ensiná-la etc. Além disso, é impossível negar que a gramática tradicional é o repositório de importantes reflexões de filósofos e filólogos — por baixo da pesada ideologia prescritiva existem interessantes sugestões de análise, além de descobertas importantíssimas sobre o funcionamento da linguagem humana em geral e das línguas em particular.

Para a pessoa que ensina português brasileiro e considera que sua formação foi, tem sido ou é insatisfatória nos dois pontos mencionados acima, é imprescindível procurar preencher essas lacunas por meio de leituras.

Para conhecer bem a doutrina gramatical tradicional, sugiro que você leia, de ponta a ponta, uma gramática de autor respeitado e respeitável, como essa:

Nova gramática do português contemporâneo, de Celso Cunha e L. F. Lindley Cintra (Rio de Janeiro: Lexikon, 2008).

Embora venha assinada por dois autores, um brasileiro (Cunha) e um português (Cintra), essa gramática de fato é uma revisão e ampliação de uma obra anterior de Celso Cunha, uma vez que Lindley Cintra ficou responsável unicamente pelo capítulo que trata da distribuição do português no mundo e da dialetologia do português

europeu. A leitura é agradável porque Celso Cunha sempre escreveu com clareza e objetividade, sem o estilo muitas vezes arrevezado e pomposo dos filólogos de sua geração (o linguista francês Paul Teyssier dizia que Celso Cunha era o único gramático brasileiro que não tinha "a pena gordurosa").

Depois dessa obra, também é importante ler e estudar a

Moderna gramática portuguesa, de Evanildo Bechara (Rio de Janeiro: Nova Fronteira, 2009).

A obra de Bechara representa um importante movimento de transição da tradição prescritiva para uma abordagem mais sintonizada com os avanços da linguística científica. Sua filiação à tradição se comprova pelo recurso às abonações literárias de autores selecionados de um grupo bastante restrito (não inclui, por exemplo, José Saramago, Clarice Lispector, Cecília Meireles, Jorge Amado, Rubem Braga, entre outros) e por diversas recaídas em posturas marcadamente prescritivas [▶527]. No entanto, a abordagem dos fenômenos linguísticos se faz muitas vezes em declarada contraposição à tradição e com base em postulados teóricos de correntes científicas contemporâneas.

Por fim, seria altamente recomendável empreender um estudo bem detido de mais uma obra:

Gramática Houaiss da língua portuguesa, de José Carlos de Azeredo (São Paulo: Publifolha, 2008).

Aqui não estamos mais no terreno da tradição gramatical, pois o autor não recorre a abonações literárias, faz uma abordagem muito bem fundamentada das questões de variação linguística e trata dos conteúdos gramaticais dentro de uma perspectiva inovadora, mas perfeitamente acessível a quem estiver familiarizado com essa área de conhecimento. Além disso, apesar do título, se concentra no que é próprio do português brasileiro. Para quem não tiver condições de ler as duas obras anteriores, recomendo que se concentre então nessa última.

Para quem quiser se inteirar das principais escolas teóricas da linguística moderna, será de grande proveito a leitura e o estudo do

Manual de linguística, organizado por Mário Eduardo Martelotta (São Paulo: Contexto, 2008).

Depois dessa primeira visita, vale a pena se aprofundar em cada um dos domínios específicos da ciência linguística lendo

Introdução à linguística, 3 volumes organizados por Anna Christina Bentes e Fernanda Mussalim (São Paulo: Cortez, 2001).

Essas leituras criarão uma base importante para a compreensão do muito do que está em jogo na prática da educação em língua materna. Sempre lamento muito quando, em eventos e debates com professores pelo Brasil afora, vejo que é difícil avançar nas discussões quando faltam algumas informações básicas, primárias, que deveriam ter sido supridas desde o início nos cursos de formação docente.

Por que esse livro

Só tomei a iniciativa de produzir esse livro depois que foram lançadas no mercado brasileiro as primeiras obras explicitamente voltadas para a descrição e a análise da nossa língua, o português brasileiro. Me refiro aos dois títulos abaixo:

Gramática do português brasileiro, de Mário A. Perini (São Paulo: Parábola Editorial, 2010).
Nova gramática do português brasileiro, de Ataliba T. de Castilho (São Paulo: Contexto, 2010).

Esses dois lançamentos representam uma radical mudança na história das publicações gramaticais brasileiras[2]. Além de trazer no título o nome da nossa língua, tal como vem sendo usado nas pesquisas mais avançadas sobre a realidade linguística brasileira, essas duas obras rompem com a tradição de vincular sempre o estudo gramatical da língua da maioria dos brasileiros à comparação entre essa língua e a língua dos portugueses (e sempre sob a perspectiva da escrita literária, como se sabe).

Em seu livro, Perini se dedica ao exame de importantes construções sintáticas da língua, a algumas classes de palavras e a alguns aspectos da fonologia. Não é, portanto, uma gramática completa, como estamos acostumados a ver. O importante é que ali não encontramos referências ao padrão literário nem comparações entre o português brasileiro e o português europeu.

A obra de Castilho, bem mais volumosa, abrange vários terrenos de investigação: a história da língua e de sua implantação em terras brasileiras, um importante debate sobre teorias linguísticas e concepção de língua, fonologia, morfologia e muita mor-

[2] Uma obra igualmente original é a *Gramática de usos do português*, de Maria Helena de Moura Neves (São Paulo: Unesp, 2001). No entanto, apesar de seus inúmeros méritos, seu *corpus* de análise é constituído exclusivamente de língua escrita (literária, jornalística, ensaística), em gêneros textuais mais monitorados, nos quais ainda persistem, ao lado das tendências inovadoras próprias do português brasileiro, usos mais conservadores, decorrentes da própria seleção textual do *corpus*. Por outro lado, a *Moderna gramática brasileira*, de Celso Pedro Luft (São Paulo: Globo, 2003), apesar do título, é essencialmente uma gramática da "língua portuguesa" no mesmo espírito das tradicionais. Já a *Gramática do brasileiro*, de C. Ferrarezi e I. M. Teles (São Paulo: Globo, 2008) não cumpre o que promete no título.

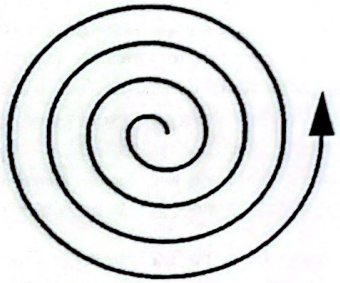
fossintaxe, semântica e pragmática. A gramática de Castilho é essencialmente uma descrição do português brasileiro urbano culto falado, pois representa uma síntese das várias décadas de trabalho do autor com o *corpus* do projeto NURC (Norma Urbana Culta), composto exclusivamente de língua falada por brasileiros de cinco grandes cidades (Recife, Salvador, Rio de Janeiro, São Paulo e Porto Alegre). Por sua abrangência e pela riqueza de seu conteúdo, a gramática de Castilho é, desde já, um clássico da literatura linguística brasileira.

Essas duas obras se destinam fundamentalmente ao público universitário e não têm pretensões didático-pedagógicas, no sentido de contemplar a comunidade escolar: professoras e professores, alunas e alunos e demais agentes educacionais. Por isso, conhecendo as demandas e expectativas do corpo docente, assim como as difíceis condições de trabalho da nossa escola pública, decidi produzir este livro, como um auxiliar para a tarefa de promoção da *reflexão sobre a língua e a linguagem* em sala de aula.

Essa tarefa bem que poderia ser cumprida pelos livros didáticos, especialmente pelos que são adquiridos pelo Ministério da Educação e distribuídos para as escolas, aos milhões de exemplares e a um custo elevadíssimo, através do PNLD (Programa Nacional do Livro Didático). No entanto, e muito infelizmente, a grande maioria desses livros (75% segundo o Guia do PNLD de Língua Portuguesa, edição 2008) se limita a querer transmitir, intacta e sem crítica, a tradição gramatical prescritiva, com sua profusa e confusa nomenclatura (encontrei, por exemplo, mais de 250 termos gramaticais diferentes numa única coleção de 5ª a 8ª séries), junto com uma concepção de "língua certa" extremamente estreita e sem correspondência com a realidade da atividade linguística dos brasileiros, tanto em suas práticas orais quanto em suas práticas escritas, inclusive dos brasileiros mais letrados, inclusive dos escritores modernos e contemporâneos (os resultados dessa pesquisa se encontram em Bagno, 2011).

Aqui a professora e o professor vão encontrar a descrição de aspectos essenciais da gramática do português brasileiro, com vasta exemplificação de usos autênticos contemporâneos, junto com propostas de atividades práticas para levar seus aprendizes a conhecer melhor o funcionamento da língua que falam e escrevem e para se apoderar do que é um **português brasileiro contemporâneo urbano culto**, que nada tem que ver com o modelo muito idealizado de "língua certa" que as gramáticas prescritivas, os livros didáticos e os meios de comunicação (através do que chamo de *comandos paragramaticais*) ainda insistem em divulgar, sem se dar conta de que aquela há muito tempo já deixou de ser a língua da maioria dos brasileiros, incluindo os da elite mais letrada, para não falar da nossa melhor literatura contemporânea.

Nesse livro aqui — faço questão de repetir — vou me valer de uma estratégia argumentativa clara e definida: todos os usos que já são característicos do português brasileiro contemporâneo, principalmente em suas *variedades urbanas de prestígio* e, mais ainda, nos *gêneros textuais escritos mais monitorados*, vão ser tomados pacificamente como legítimos e indiscutíveis, por mais que os agentes do prescritivismo purista se escandalizem com isso. O argumento para justificar essa postura é muito simples: esses usos representam *mudanças linguísticas já devidamente implantadas no nosso sistema linguístico* e, como todas as mudanças desse tipo, são usos incontornáveis, inevitáveis, irreprimíveis, inconscientes (porque já perfeitamente integrados à nossa cognição da língua) e irretrocessíveis, pois é sabido que a mudança linguística se faz sempre para adiante, num processo cíclico, mas em forma de uma espiral que retorna a fases anteriores, sempre tomando distância, porém, com relação aos pontos já percorridos:



Contra a mudança linguística não há nada que se possa fazer: ela é inevitável e é da própria natureza das línguas, tal como a mudança é da natureza de tudo o que existe na sociedade, no mundo e no universo. Em vez de lutar contra ela — o que é sempre uma guerra perdida de antemão —, mais vale buscar formas de convívio sadio e tranquilo com ela. E é o que vamos tentar fazer aqui. Não há nada que se possa fazer contra a mudança linguística porque ela é impulsionada pelos próprios falantes. É no mínimo irônico ver a tradição purista tentando vencer o invencível: o processamento sociocognitivo da língua por seus próprios falantes nativos, incluídos os puristas!

Parodiando o velho Marx, os linguistas têm se limitado a *descrever* as relações entre língua e sociedade: trata-se, agora, de *transformá-las*. Essa transformação pode ser feita por meio da formulação de um "discurso herético", como escreve Pierre Bourdieu (1996: 118), destinado a produzir um "novo senso comum". Por isso o qualificativo de pedagógica que aparece no título desta gramática — ela descreve, mas também propõe uma nova norma linguística para o ensino, uma norma que não é extraída do nada nem dos gostos pessoais do autor (como é frequentemente o caso com as gramáticas prescritivas), mas uma norma que já existe, "tácita ou recalcada", e que tem de ser legitimada para que o Brasil exorcize de vez o fantasma colonial que ainda assombra nossas concepções de língua e de língua e ensino.

O QUE ENSINAR NA ESCOLA

Se a escola tem a função de *ensinar*, é (ou deveria ser) claro, óbvio, nítido e transparente que os objetos do ensino devem ser coisas que as pessoas ainda não sabem. Se procuro uma autoescola, é porque não sei dirigir um carro, ou porque o que sei não é suficiente. Se procuro uma escola de língua alemã, é porque não sei alemão, ou porque o que sei não é suficiente. Então, se quem vai para a escola, no Brasil, são brasileiras e brasileiros falantes do português brasileiro, por que é necessário *ensinar* essa língua a essas pessoas?

A resposta está no fato de que por trás do rótulo genérico língua se esconde um universo multidimensional de saberes, práticas, crenças, histórias, ritos, instrumentos, mecanismos de poder, papéis sociais, tradições culturais, para dizer o mínimo [▶55]. Boa parte disso tudo a gente aprende em casa, com a família, na nossa comunidade, nos grupos que fazemos parte, nas redes sociais que nos movimentamos — tal como aprendemos em casa a não usar preposição antes de pronome relativo, conforme acabo de fazer [▶901]. Mas outro tanto de conhecimento só nos é revelado pelas instituições às quais a sociedade tem atribuído precisamente o papel de transmitir esse legado. Justamente por isso é que essas instituições — que vamos aqui chamar pelo nome singular de escola — precisam ser *democráticas e democratizadoras*, respeitosas da pluralidade cultural e da diversidade humana que compõe a sociedade.

A escola é, antes de mais nada, lugar de *socialização*. Ao entrar na escola, cada aprendiz entra também em contato com outras comunidades, outros modos de ser, outros jeitos de falar, outras redes sociais. Ao mesmo tempo, a escola é a agência que estabelece a relação entre cada aprendiz e o Estado, responsável pela educação de seus cidadãos, mesmo que se trate de escolas privadas. Assim, se o Estado quer ser democrático, é imprescindível que ofereça aos cidadãos uma escola democrática.

Além desses aspectos, a escola também é a principal agência de *letramento* dos aprendizes. Certamente estamos aqui lidando com uma noção restrita de letramento, o *letramento escolar*, uma vez que o próprio aprendizado da língua materna desde a primeira infância constitui um processo de letramento: a aquisição da gramática da língua materna, a formação de uma intuição linguística, a capacidade de reconhecer uma manifestação linguística como bem ou malformada, como pertencente ou não à própria língua — tudo isso também é letramento. E mesmo que a gente se refira a letramento como a aprendizagem da leitura e da escrita, tem pessoas que não precisaram frequentar escola para aprender a ler e a escrever. No entanto, o número dessas pessoas é muito pequeno quando comparado com a multidão daquelas que só vão fazer essa aprendizagem no ambiente escolar.

Se todo ser humano dotado de suas plenas capacidades sociocognitivas e físicas é capaz de *falar* uma ou várias línguas, já *ler e escrever* é um conhecimento que exige a ativação de outros mecanismos cognitivos, sociais e culturais. Falar é tão natural para um ser humano quanto respirar, ver, ouvir, sentir odores, caminhar em posição ereta, sonhar à noite etc. Ler e escrever, por sua vez, é da mesma ordem de coisas de tocar piano, dirigir um automóvel, tricotar um xale, pintar um quadro, construir uma casa etc. São práticas socioculturais que exigem um aprendizado relativamente longo e contínuo para que o aprendiz se apodere delas a ponto de exercê-las com habilidade e destreza. Sendo assim, o ensino da leitura e da escrita, e o acesso aos discursos sociais que se valem delas, é a tarefa primordial da educação em língua materna na escola.

Muitos pesquisadores têm postulado, no entanto, que os seres humanos não somente *falam* línguas como também *falam sobre* as línguas e, principalmente, falam sobre a língua que falam. É da nossa própria natureza sociocultural a reflexão *epilinguística*, isto é, tomar a língua como objeto de apreciação, depreciação, avaliação (positiva ou negativa), crítica, prazer estético, curiosidade, conhecimento do mundo etc. Por isso, também cabe à escola sistematizar essa propensão a falar sobre a língua, de modo a favorecer, com ela, mais uma vez, o processo de letramento.

Ler, escrever e refletir sobre a língua. Essas três tarefas — que no fundo são uma só: *desenvolver o letramento* — constituem toda a missão da escola no que diz respeito à educação em língua materna. Não há tempo a perder com outras práticas que já se comprovaram absolutamente irrelevantes e inúteis para se cumprir essa missão:

- Decorar uma nomenclatura gramatical profusa, confusa e muitas vezes incoerente e inconsistente? — Não!
- Classificar uma palavra solta, sem texto, cotexto nem contexto real de uso? — Não!
- Identificar numa frase uma categoria gramatical sem atentar para os efeitos discursivos que ela permite produzir? — Não!
- Proceder à análise sintática de uma oração apenas para rotular seus elementos constituintes? — Não!
- Desconsiderar um texto em sua rede de significações e de sentidos para nele pinçar apenas palavras de uma determinada classe gramatical? — Não!
- Produzir "redações" com temas irrelevantes, sem definição de tipo nem de gênero textual? — Não!
- Ser convencido de que só existe uma maneira correta de dizer ou escrever e que todos os demais usos da língua são errados e feios? — Não!

Nenhuma dessas práticas garante uma educação linguística honesta, relevante e útil para a formação cidadã.

Por isso, volto a insistir, essa gramática aqui é para quem ensina, para que as pessoas responsáveis pela promoção do letramento de outras pessoas conheçam a fundo seu objeto de trabalho. Não se deve ensinar gramática na escola, mas quem ensina na escola deve conhecer muitíssimo bem a gramática!

Na verdade, pode ser uma contradição em termos dizer que *não se deve ensinar gramática* se o objetivo é *promover o letramento*. Mas não é uma contradição em termos quando tomamos o termo *gramática* em sua acepção mais corriqueira e reducionista: *a nomenclatura gramatical*. Saber o que é uma oração subordinada substantiva objetiva direta reduzida de infinitivo *não é saber gramática*: é saber aplicar um *rótulo* a um retalho de linguagem.

Eu posso encontrar uma lataria de carro enferrujada, abandonada no meio do mato, e identificar a marca, o modelo, o ano e talvez até a cor do veículo. Mas não posso saber a quem ele pertenceu, aonde sua dona ou seu dono foi com ele, quem mais viajou nele, se aquele carro foi multado alguma vez, se esteve envolvido em algum acidente... enfim, não vou saber nada da *história* do carro.

O mesmo acontece quando se isola uma frase de um texto ou — tanto faz, o dano é idêntico — quando se inventa inteiramente uma frase para que seja submetida a uma análise morfossintática. Alguém pode classificá-la com base em sua estrutura aparente, mas nunca terá como conhecer a *história* daquela frase — e só existe *linguagem*, atividade humana, onde existe *história*, vida humana.

Saber gramática é muito mais do que rotular. Saber gramática é algo tão entranhado em cada pessoa que é simplesmente impossível falar, ouvir, ler, escrever ou refletir sobre a língua sem ativar esse conhecimento gramatical intuitivo e poderoso. Vamos ver um exemplo?

Num livro didático de 5ª série, encontrei a seguinte atividade, proposta depois da leitura de um texto que narrava o mito de Teseu, Ariadne e o Minotauro:

> 5 Leia as frases a seguir e, no caderno, explique a diferença de sentido entre as palavras ou expressões destacadas.
> (a) "Teseu, cauteloso, para e vigia os mínimos esconderijos [...]"
> (b) Ariadne pensou até mesmo nos aspectos mínimos.
> (c) "[...] eu já de longe fico sabendo que você está vivo."
> (d) "O rei Minos nem de longe imaginava o plano de Teseu."
> (e) "[...] o velho Egeu não teve um único momento de repouso."
> (f) "Aquele foi um momento único para Teseu."
>
> 6 Observe:
> "Sob seu sábio governo, a Grécia conheceu a paz. E Atenas, a prosperidade."
> (a) Que palavra a vírgula substituiu na segunda frase? Por que o autor usou esse recurso?
> (b) Crie um exemplo semelhante, usando a mesma pontuação. [Sarmento, 2008, vol. 5: 17-18]

Observe que a autora não pede nenhuma classificação, não quer que o usuário do livro rotule nada, mas que depreenda "a diferença de sentido" provocada pelo uso de uma mesma palavra ou locução em determinados trechos do texto lido e em frases que, embora inventadas, mantêm relação com a narrativa e permitem que o aprendiz se conscientize daquelas diferenças. O mesmo vale para a segunda atividade, em que não se solicita o emprego de rótulos, mas a depreensão dos valores de um sinal de pontuação e de como ele pode conferir ritmo e concisão à produção textual, além de garantir a *coesão* textual.

Fazer essas reflexões sobre os usos da língua, reconhecer e apoderar-se dos efeitos de sentido obtidos pelo emprego consciente dos recursos que a língua oferece — isso sim é *estudar gramática!*

Portanto, não tem cabimento a pergunta, tantas vezes repetida, "é ou não é para ensinar gramática?", já que a resposta é *sim* ou *não*, dependendo da concepção de gramática que se tem em mente: **sim, sim e sim** para a gramática como depreensão dos sentidos de um texto e dos mecanismos linguísticos que permitem a produção desses sentidos; **não, não e não** para a rotulação estéril, a classificação mecânica, a decoreba que ofende a inteligência.

O NÃO ENSINO DA NORMA-PADRÃO

Outra pergunta resultante de muitos mal-entendidos tem a ver com o "ensino da norma-padrão". O que se entende por *norma-padrão*, nos estudos mais recentes sobre variação linguística e ensino, é o modelo de língua descrito-prescrito pela tradição gramatical, uma língua extremamente idealizada, construída com base nos usos de um grupo não muito amplo de escritores e, mesmo assim, não de todos esses usos, mas só daqueles que o próprio gramático considera exemplares ou recomendáveis. Essa norma-padrão — *escrita, literária e obsoleta* — é, por isso mesmo, repleta de arcaísmos, de fósseis linguísticos, de regras que vão contra a intuição gramatical de qualquer falante da língua. Como se não bastasse, ela é inevitavelmente *anacrônica*, porque recorre a um cânone literário do passado, de modo que nem sequer na literatura viva, contemporânea, é possível reconhecer o uso integral do que ela prescreve.

Diante disso, a resposta é clara: **não se deve ensinar norma-padrão na escola.** A educação em língua materna não é sinônima de um ensino exclusivo de uma única modalidade de emprego da língua, muito menos de uma modalidade obsoleta e anti-intuitiva[3]. Educar em língua materna é permitir o acesso dos aprendizes **ao maior número possível de modalidades faladas e escritas de sua língua**, modalidades que só se realizam empiricamente, concretamente, na forma de gêneros textuais.

3 Esse é o equívoco fundamental da proposta de L. C. Assis Rocha em seu livro *Gramática nunca mais!* — propor o ensino exclusivo de uma norma obsoleta, que o próprio autor, aliás, deixa de obedecer em diversos trechos de seu livro... Outro equívoco é propor o ensino dessa norma-padrão arcaica por meio de exercícios estruturais de preenchimento de lacunas, uma metodologia de ensino ultrapassada e descartada pela pedagogia de línguas há muitas décadas.

É no trabalho com os mais variados gêneros textuais — falados e escritos — que os aprendizes tomarão consciência da multiplicidade de usos possíveis da língua. Um folheto de cartomante distribuído na rua merece ser estudado e analisado em sua constituição textual: seu vocabulário, suas estratégias argumentativas, seus recursos morfossintáticos, tanto quanto um editorial de uma revista de grande circulação. É evidente que *a literatura* tem de se destacar na educação em língua materna, mas não para servir de "modelo a imitar", e sim pelo que representa de fruição estética e de elemento constitutivo da cultura e da identidade de um indivíduo e da coletividade.

Não tem cabimento, por exemplo, apresentar o paradigma da conjugação verbal com as seis pessoas do português clássico — *eu, tu, ele(a), nós, vós, eles(as)* —, já que essa conjugação não corresponde a absolutamente nenhum uso real de nenhuma das variedades do português brasileiro falado ou escrito, nem do português europeu, angolano, moçambicano etc. É imperioso que se apresente os diferentes paradigmas verbais em vigor no português brasileiro contemporâneo, com *você*, com *a gente*, com *tu foste* e com *tu foi* etc., porque são esses paradigmas variáveis que de fato estão vivos na nossa sociedade. É um crime pedagógico esconder a realidade da língua aos que procuram a escola precisamente para conhecer essa realidade!

É claro que no trabalho com um texto em que apareça uma forma verbal com o índice de pessoa *vós* a professora terá uma excelente oportunidade para mostrar aos alunos que a língua muda com o tempo, que faz mais de trezentos anos que ninguém usa o índice de pessoa *vós*, que ele só aparece em determinados textos — antigos, religiosos ou humorísticos, por exemplo. Os alunos têm todo o direito de conhecer o índice de pessoa *vós* e outros arcaísmos, tanto quanto têm todo o direito de saber o que significa *sine qua non, grosso modo, data venia* ou *errare humanum est*, o que significa *yo no creo en brujas pero que las hay, las hay*, o que significa *l'État, c'est moi* ou *noblesse oblige*, o que significa *last but not least*, sem que para tanto precise conhecer a fundo latim, espanhol, francês ou inglês.

As formas linguísticas obsoletas, arcaicas ou em vias de extinção não precisam ser *ensinadas* para que os alunos entrem em contato com elas. Quando e se aparecerem em textos autênticos, que mereçam ser lidos e estudados, a professora pode explicar do que se trata, sem impor aos alunos um conhecimento ativo daqueles fósseis. Sem propor uma atividade de passar frases com o índice de pessoa *tu* para o índice de pessoa *vós* fazendo as devidas concordâncias — devidas a quem?

Não existe vida social sem que se estabeleça *normas* para a conduta em sociedade. Podem ser normas ditadas de cima para baixo, impostas ao conjunto das pessoas por alguma instituição hierarquicamente superior, ou podem ser normas surgidas do próprio convívio social, da própria atividade sociocultural e sociopolítica da comunidade. É isso que faz a diferença entre o *normativo* (de cima para baixo) e o *normal* (surgido e difundido no seio da comunidade). Na história das sociedades, é muito frequente que o *normal*, de tão aceito e praticado, seja regulamentado e *normatizado* pelas instâncias superiores, como a legislação. Durante séculos e séculos, o casamento foi indissolúvel nas sociedades ocidentais, cristãs. Com o passar do tempo, no entanto, foi se tornando *normal* a separação dos casais, mesmo que, nas aparências, permanecessem casados. A pressão dessa normalidade foi tamanha que o próprio Estado reconheceu o direito à separação, assim como muitas legislações vêm reconhecendo os direitos das mulheres, das minorias étnicas, das crianças, dos homossexuais, dos refugiados, dos deficientes físicos etc.

O mesmo ocorre na língua. A pressão da mudança — que se processa na sociedade, impulsionada pelos falantes em suas interações — acaba por transformar uma forma linguística inovadora num uso tão *normal* que, mesmo enfrentando a reação de uma minoria (os gramáticos mais prescritivistas, os puristas, os reacionários em geral), acaba por se impor ao conjunto da sociedade. Vamos tratar mais demoradamente desses processos de mudança no capítulo 4.

A *norma-padrão* tradicional acaba perdendo espaço para a *norma real, habitual, normal*, pelos usos feitos pelos falantes em suas atividades linguísticas cotidianas. É dessa norma real, habitual, normal, que vamos tratar nesse livro. Não adianta os prescritivistas afirmarem que eu deveria escrever *"vamos tratar* neste *livro"*, porque no vernáculo geral brasileiro o demonstrativo *este* se extinguiu e, quando usado, não segue nem de longe as determinações impostas pela tradição prescritiva.

Sendo essa uma gramática do português brasileiro, o que nela vai aparecer são as formas genuinamente brasileiras de falar e de escrever. E uma vez que a função da escola é ensinar o que as pessoas não sabem, é nos *gêneros textuais escritos mais monitorados* que vamos buscar nossas exemplificações do real funcionamento da gramática do português brasileiro. Porque são esses gêneros escritos mais monitorados que é dever da escola levar seus alunos a conhecer, reconhecer, dominar e empregar. O recurso a esses gêneros serve também como argumento contra aqueles que tentam não parecer autoritários ou puristas alegando que "na fala, tudo bem, mas, na escrita, nunca", como se fala e escrita fossem duas entidades totalmente diferentes, estanques e isoladas.

Nessa gramática, vamos comparar sempre o VGB (vernáculo geral brasileiro) com a TGP (tradição gramatical do português), dando sempre ênfase e prioridade político-pedagógica ao VGB. Com isso, estamos assumindo a postura, igualmente política, de legitimar no ensino os usos mais difundidos no VGB, de forma a abandonar a arcaica separação entre "certo" e "errado". Não sugerimos que o aluno não tenha acesso à norma-padrão veiculada pela TGP, até porque ela vai surgir inevitavelmente nos textos que ele vai ler em seu processo de escolarização. Só reivindicamos que ela não seja usada como mecanismo didático de negação do VGB nem como instrumento para depreciar nossa língua materna. Não há por que negar legitimidade ao que já está legitimado por todos os falantes, inclusive pelos urbanos e altamente letrados, inclusive pela nossa melhor literatura dos últimos cem anos.

LIVRO I EPISTEMOLOGIA DO PORTUGUÊS BRASILEIRO

a Luiz Antônio Marcuschi

1

a caverna implodida

por uma concepção não-platônica de língua

Todo trabalho acerca da linguagem e das línguas, para ser coerente do ponto de vista teórico, político-ideológico, pedagógico e metodológico, tem de definir o que entende por língua. Só é possível falar da língua depois de definir o que se entende por esse termo, uma vez que a língua não está diretamente acessível como um objeto concreto semelhante às pedras, às plantas, aos animais, às substâncias químicas etc., que constituem

os objetos de outras ciências. Nesse capítulo, vamos criticar o que chamo de platonismo linguístico (uma filosofia dualista), passar em revista as correntes teóricas que o sustentaram e ainda o sustentam, e propor uma concepção alternativa — monista — da linguagem e das línguas.

Essa concepção, evidentemente, não é a "*melhor*" nem a mais "verdadeira" — é simplesmente uma questão de *escolhas* filosóficas e ideológicas, sempre imbuídas do subjetivismo e das crenças pessoais do postulante. A verdade sobre a língua — e sobre qualquer outra coisa — está desde sempre e para sempre inacessível à mente humana por causa de suas restrições naturais, de suas limitações cognitivas (por mais extensas que as consideremos!). Mas isso não é impedimento — ao contrário, é o estímulo — para a busca incessante de conhecimento.

1.1 Platão

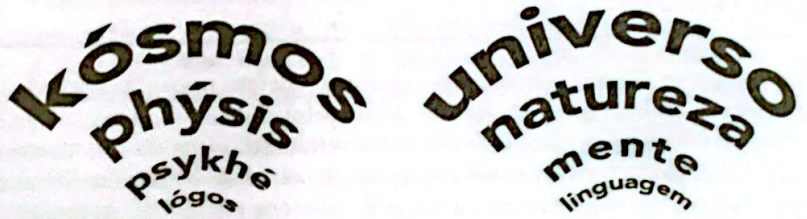
O filósofo grego Platão (428/427 — 348/347 a.C.) é um dos nomes mais importantes de toda a cultura ocidental. O impacto de seu pensamento é tamanho, sua influência é tão profunda que o filósofo inglês Alfred North Whitehead (1861-1947) chegou a dizer que "a tradição filosófica europeia é uma série de notas de rodapé a Platão". De fato, o filósofo grego levantou os principais problemas e temas que a filosofia e outras áreas de conhecimento querem resolver e investigar até hoje. Sem Platão, não só a filosofia europeia seria diferente, mas também toda a tradição teológica cristã. Sua contribuição para os estudos da linguagem também é inegável, e dela trataremos no capítulo 9.

Não podemos desdobrar aqui todas as múltiplas facetas do sistema filosófico platônico. O que vai nos interessar nesse capítulo é o seu dualismo característico. Uma filosofia dualista é aquela que repousa numa dupla de conceitos fundamentais que se opõem de modo inconciliável. Na filosofia de Platão, a oposição fundamental é entre o mundo sensível (*fenomênico*), aquele que pode ser apreendido pelos nossos sentidos — portanto, o mundo corpóreo, material, físico —, e o mundo cognoscível (*numênico*), aquele que só pode ser apreendido pela nossa inteligência — portanto, o mundo mental, espiritual, metafísico. Nesse sistema filosófico, o conhecimento fornecido pelos sentidos humanos é chamado de "ilusão" (a famosa "ilusão dos sentidos"), porque não nos permite acesso direto à verdadeira realidade das coisas, à verdade, enfim. Só o exercício da inteligência, do raciocínio profundo e (nas doutrinas religiosas) da contemplação mística permite que nos libertemos dessa ilusão dos sentidos e tenhamos acesso à essência real das coisas (e/ou à plenitude espiritual).

Nessa visão platônica das coisas inteligíveis, a linguagem ocupa o lugar mais baixo da hierarquia dos objetos que dão acesso ao conhecimento. De fato, os filósofos

gregos acreditavam que o entendimento da organização da linguagem (*lógos*) permitiria o acesso à organização da mente (*psykhé*) e que, por sua vez, a compreensão do funcionamento da mente poderia levar a um entendimento da organização do mundo natural (*phýsis*) e, daí, à organização do próprio universo (*kósmos* — palavra que significa, originalmente, "harmonia", "ordem", em oposição a *kháos*, o "caos", a desordem, o vazio).

Haveria portanto, entre esses diferentes níveis de realidade, uma relação especular (do latim *spéculu-*, "espelho"), em que a organização e o funcionamento de cada um dos objetos investigados seria um reflexo da organização e do funcionamento dos demais objetos, numa gradação hierárquica:



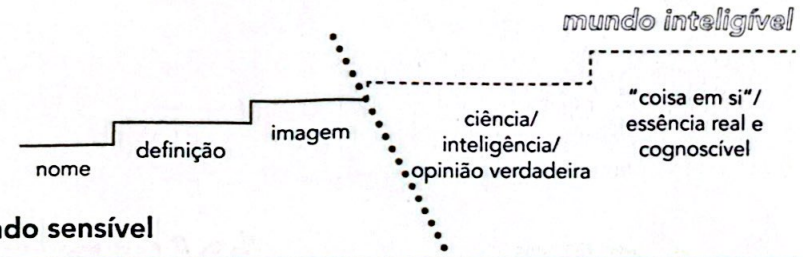
Por ser o mais imediatamente acessível, o *lógos*, o discurso (linguagem portadora de razão), será o ponto de partida da investigação filosófica. Para nós, estudiosos apaixonados da linguagem, dá tristeza ver nosso objeto de estudo ocupar a parte menos nobre da hierarquia dos filósofos gregos. Realmente, no que diz respeito a Platão, por exemplo, como nos explica Neves (2002: 19-20),

> a linguagem não é o fim último da investigação platônica. Pelo contrário, na filosofia platônica, a linguagem é considerada apenas como um fraco auxiliar para se chegar ao autó, isto é, à própria 'coisa', ao objeto verdadeiramente real e cognoscível. Entre os cinco elementos que tornam possível o conhecimento (...), o nome ocupa o degrau inferior, e o lógos (que é a definição, a proposição), o degrau imediatamente seguinte. Por eles (isto é, pela linguagem) se chega à imagem (o eídolon), terceiro degrau, e só então se sobe à ciência (epistéme), à inteligência (noûs) e à opinião verdadeira (alethés dóxa), que estão na alma, não possuem modo sensível de existência e, por isso, escapam ao mutável (que está no nome, que está no lógos, que está na imagem); aí beiram a essência imutável, verdadeira, estável, sempre idêntica a si mesma, que é a própria coisa (o autó). Isso quer dizer que a linguagem já supõe a existência das coisas, e nem nome, nem lógos, nem imagem têm nenhuma fixidez. (...) E a linguagem é, pois, algo inferior.

É o que podemos ler nas palavras do próprio filósofo (*Carta VII*, 342e):

> Os quatro [fatores que conduzem ao conhecimento] manifestam tanto a qualidade quanto o ser de cada coisa *por meio da debilidade da linguagem* [*dia to tōn logōn asthenes*] [grifo meu].

Relegar a linguagem ao plano mais baixo, como o mais fraco dos instrumentos que permitem o acesso à essência verdadeira das coisas, é perfeitamente coerente com os princípios da filosofia platônica, onde o conhecimento mais importante é aquele que não pode ser apreendido pelos sentidos, mas somente pela inteligência.



Os três primeiros degraus se encontram no mundo concreto, no mundo sensível-sensual-sensorial, daquilo que pode ser apreendido pelos sentidos — ou seja, o mundo dos fenômenos (do grego *phainómenon*, termo derivado do verbo *phaíno*, "aparecer, revelar-se, ser perceptível"). Os dois últimos se encontram no mundo inteligível, do espírito, ao qual só se tem acesso por meio do exercício intelectual, do raciocínio, da abstração, e onde habitam as célebres Formas *Ideais* da filosofia platônica — ou seja, o mundo dos números (do grego *noúmenon*, particípio presente do verbo *noéo*, "pensar", "raciocinar"), que o filósofo alemão Immanuel Kant (1724-1804) vai chamar de "coisa em si" (*das Ding an sich*). Para Platão, a realidade sensível é como a imagem, a figura, que representa, de modo imperfeito, a verdadeira essência dos objetos. É preciso ultrapassar essa cópia, essa imitação (*mimésis*), e buscar o ser autêntico das coisas. Daí a condenação que Platão faz da poesia: ela não diz a realidade das coisas, apenas a transforma em alegorias, símbolos, metáforas, por meio da linguagem que já é, em si mesma, uma cópia imperfeita do Real.

É fácil compreender por que os primeiros grandes teólogos do cristianismo se valeram amplamente da filosofia platônica para falar da diferença entre o reino terrestre e o reino celeste, dos atributos humanos e dos atributos divinos, entre o corpo mortal e a alma imortal etc. Já nos primórdios do cristianismo, o apóstolo Paulo, na conclusão de seu belo "Hino ao amor" (1 Coríntios 1,11-13), retomava a noção platônica da imagem imperfeita, representada no espelho, cuja visão distorcida seria corrigida pelo acesso à graça divina, obtida por meio da fé, da esperança e do amor:

> 12 βλέπομεν γὰρ ἄρτι δι' ἐσόπτρου ἐν αἰνίγματι, τότε δὲ πρόσωπον πρὸς πρόσωπον· ἄρτι γινώσκω ἐκ μέρους, τότε δὲ ἐπιγνώσομαι καθὼς καὶ ἐπεγνώσθην.
>
> 12 Videmus enim nunc per speculum in aenigmate,
> tunc autem facie ad faciem; nunc cognosco ex parte,
> tunc autem cognoscam, sicut et cognitus sum.

> 12 Agora, vemos em espelho, de modo enigmático;
> mas então será face a face; agora conheço só em parte;
> mas então conhecerei como sou conhecido.

(Convém lembrar, no entanto, que na época de Paulo o espelho não era feito de vidro revestido de tinta metálica, como hoje, o que permite uma imagem refletida perfeitamente idêntica ao "original". Na antiguidade, o espelho era, em geral, um círculo de metal brilhante — como o ouro ou a prata —, liso e polido, onde as pessoas podiam se mirar, mas sem obter uma imagem perfeita. Daí a equiparação de "espelho" com "imagem distorcida", presente no platonismo e no discurso paulino).

O importante filósofo-teólogo que foi santo Agostinho (354-430) deixa transparecer sua filiação platônica no livro intitulado *A Cidade de Deus (De Civitate Dei)*, em que opõe a "cidade dos homens", material, imperfeita, impura, pecadora, transitória etc., à "cidade de Deus", transcendente, pura, perfeita, imaculada, eterna etc. (e isso muito embora ele afirmasse, após sua conversão, ter abandonado a "filosofia pagã").

Numa de suas obras mais divulgadas, *A República*, Platão vai elaborar uma síntese de sua filosofia, síntese que se tornará um dos textos mais influentes de toda a história do Ocidente. É o famosíssimo mito da caverna, que vamos resumir a seguir:

> Imagine uma grande caverna, ligada ao mundo exterior por uma passagem longa o suficiente para impedir que qualquer luz do dia penetre na própria caverna. Olhando para a parede do fundo, com as costas voltadas para a entrada, está uma fila de prisioneiros. Não só seus membros estão acorrentados como também têm seus pescoços presos, de maneira que não conseguem mover as cabeças e, portanto, não podem olhar um para o outro, nem mesmo podem ver nenhuma parte de si mesmos. Tudo o que podem ver é a parede à sua frente. Eles estão nessa situação desde sempre e não conhecem nada além dela.
>
> Na caverna, por trás deles, existe uma grande fogueira. Sem que eles saibam, existe uma mureta da altura de um homem entre eles e a fogueira e, do outro lado dessa mureta, pessoas que passam constantemente de lá para cá, levando coisas em suas cabeças. As sombras desses objetos são projetadas na parede em frente aos prisioneiros pela luz da fogueira e as vozes das pessoas que carregam os objetos ecoam sobre a mureta e chegam aos ouvidos dos prisioneiros. Assim, as únicas entidades que os prisioneiros apreendem ou experimentam em toda a sua existência são essas sombras e esses ecos. Desse modo, seria natural para eles acreditar que sombras e ecos constituem toda a realidade que existe — e seria a essa "realidade" e à experiência que eles têm dela que toda a sua conversa se referiria.
>
> Se um dos prisioneiros conseguisse soltar as correntes, estaria tão entrevado por toda uma vida de imobilidade na penumbra que só girar o corpo já seria doloroso e incômodo para ele, assim como a fogueira ofuscaria seus olhos. Ele se acharia confuso e atordoado e desejaria dar as costas de novo e encarar a parede de sombra, a realidade que ele entende. Se fosse arrastado de uma vez para fora da caverna, para o mundo da deslumbrante luz do dia, ficaria cego e desnorteado, e demoraria muito até que

> conseguisse ver ou entender qualquer coisa. Mas então, quando já estivesse habituado a viver no mundo exterior, se tivesse de retornar à caverna, de novo ficaria temporariamente cego, dessa vez por causa da escuridão. E tudo o que dissesse aos outros prisioneiros sobre suas experiências seria incompreensível para aquelas pessoas cuja linguagem só teria sombras e ecos como referência.

A conclusão é que nós, seres humanos, somos como os prisioneiros da caverna: acreditamos em sombras e ecos, porque estamos presos aos nossos sentidos e à deformação da realidade que eles suscitam em nossas mentes. O dualismo aqui é evidente: caverna/luz do dia; sensação/cognição; ilusão/realidade etc.

Muito curioso é verificar que a filosofia platônica tem uma série de pontos em comum com o budismo, doutrina religioso-filosófica elaborada por Sidarta Gotama (563?-483? a.C.), o Buda ("Iluminado"), que viveu na Índia no século imediatamente anterior a Platão. O budismo também rejeita a materialidade das coisas, promove a negação dos sentidos e das ilusões da percepção sensorial, preconiza a rejeição do corpóreo, em busca da libertação dos sentidos, da supressão do desejo e da consciência individual (*Nirvana*). Buda e Platão professam a doutrina da reencarnação (que no platonismo se chama *metempsicose*), isto é, a possibilidade de uma alma humana (imortal) voltar a ocupar um corpo mortal para prosseguir em sua busca pela inteligência absoluta e total. Se existiu alguma conexão entre Platão e o budismo é algo que dificilmente poderemos descobrir.

1.2 Platonismo na linguística: gramática tradicional

A oposição fundamental da filosofia platônica — sensível *vs.* inteligível — se traduz também na oposição entre *corpo* e *alma*. A partir daí, podemos elencar uma série quase interminável de pares opostos que têm surgido e prosperado no pensamento ocidental:

- abstrato/concreto
- consciente/inconsciente
- divino/humano
- essência/aparência
- eterno/efêmero
- filogênese/ontogênese
- forma/substância
- ideia/matéria
- imanente/transcendente
- imortal/mortal
- imutável/mutável
- mente/extensão (Descartes)
- metafísico/físico
- natureza/cultura
- objetivo/subjetivo
- racional/irracional
- racionalismo/empirismo
- razão/emoção
- realismo/nominalismo (filosofia medieval)
- sagrado/profano
- social/individual
- universal/particular
- virtual/real
- etc.

Com o advento do cristianismo e sua propagação por todo o Ocidente, cada um desses pares também recebeu uma dupla carga de avaliação: positiva (para o que é divino, racional, sagrado, eterno, universal etc.) e negativa (para o outro elemento do par). Daí para a oposição maniqueísta tradicional entre bem e mal nem foi preciso dar um passo. E, é claro, numa religião que herdou a misoginia (aversão à mulher) do judaísmo, convém também incluir o par masculino (positivo) *vs.* feminino (negativo). De fato, desde sempre no cristianismo as mulheres só tinham dois modelos nos quais basear sua existência: o da santa e o da meretriz ou, pior ainda, o da meretriz que se tornou santa (Maria Madalena, Maria Egipcíaca, Taís).

Inevitavelmente, as dicotomias de inspiração platônica também exerceram sua pesada influência sobre os estudos da linguagem. Desde o surgimento da tradição gramatical entre os filólogos alexandrinos do século III a.C. até algumas das escolas teóricas da linguística do século XX, o mito da caverna tem ressurgido — aberta ou veladamente e sob as mais diversas formas e figuras — na abordagem filosófica e científica dessa impressionante faculdade humana que é a linguagem.

A própria separação entre *linguagem* (universal) e *língua* (idioma particular), tão impregnada nas ciências linguísticas, é sintomática dessa filiação platônica.

Já na constituição da gramática tradicional, doutrina que norteia até hoje tantas discussões sobre as línguas e principalmente o seu ensino, as dicotomias se manifestaram. Para os filólogos alexandrinos, a oposição crucial se dava entre *fala* e *escrita*. Para eles, grandes admiradores da produção literária do passado glorioso da Grécia clássica, somente a *língua escrita literária* merecia atenção, análise, descrição e teorização, e somente ela poderia servir de base para a *prescrição* de normas do "bem falar" e do "bem escrever". Nessa visão, a fala era caótica e desregrada, o lugar do erro e do equívoco, enquanto a escrita (concebida como algo homogêneo) era límpida e

regulada. Ao fazer isso, eles incorreram num equívoco fundamental: a comparação entre duas entidades sociolinguísticas completamente distintas:

1. a língua falada espontânea no cotidiano da Alexandria do século III a.C.;
2. a língua escrita literária da Atenas do século V a.C.

Temos aqui uma análise distorcida, resultante de três confusões:

1. a confusão de *registro de modalidade* (fala espontânea/escrita monitorada);
2. a confusão *temporal* (século V/século III a.C.);
3. a confusão *espacial* (Atenas, Grécia, Europa/Alexandria, Egito, África).

Ora, não se pode confundir a fala espontânea com a escrita literária, monitorada. Não estamos aqui defendendo a rígida separação entre fala e escrita (que vamos negar veementemente em [▶196]), mas sim a impossibilidade de comparar *registros, estilos de uso, gêneros discursivos* completamente diferentes.

Com isso, os alexandrinos definiriam os rumos dos estudos gramaticais e da pedagogia das línguas por mais de dois mil anos. Acreditando na perfeição irrefutável da língua escrita literária, eles vão criar, com base nela, um modelo ideal de "língua correta", que deveria servir de meta a ser alcançada por todo cidadão (cidadão mesmo, no masculino, pois só os homens tinham direito à educação e à cultura letrada naqueles tempos... e em muitos lugares até hoje!). Assim, toda a língua falada e os demais usos possíveis da língua escrita foram considerados como "prisioneiros da caverna", enquanto essa *norma-padrão* idealizada (e a noção de *ideal* aqui não é gratuita) seria a "língua" em sua essência mais pura, inteligível somente pelas mentes mais brilhantes e dispostas a abandonar a escuridão. Nem é preciso dizer que aí está a origem das noções de *certo* e *errado* (mais um dualismo) que tanto estrago têm feito ao longo da história da humanidade.

1.3 Platonismo estruturalista: Saussure

Quando, no início do século XX, através da divulgação de ideias atribuídas ao linguista suíço Ferdinand de Saussure (1857-1913), foram criadas as bases da escola teórica que passaria a se chamar estruturalismo, ali também se optou pelas dicotomias, todas de inspiração nitidamente platônica, uma vez que em cada par tinha uma entidade concreta/empírica/perceptível, oposta a uma entidade abstrata/dedutível/idealizável:

LÍNGUA (*LANGUE*)	FALA (*PAROLE*)
Sistema abstrato, homogêneo, composto de todas as realizações potenciais de expressão, formado por unidades que se opõem entre si, compartilhado por todos os falantes de um grupo social; é invariável pois não pode ser alterado por nenhum falante individual.	Manifestações individuais, particulares, heterogêneas, das potencialidades do sistema; apresenta variação, mas essa variação não prejudica a intercompreensão dos falantes, uma vez que todos eles reconhecem que, por trás das formas variantes, existe, num nível mais abstrato, uma forma comum, geral.
SIGNIFICADO	**SIGNIFICANTE**
Conceito abstrato, ideia na mente, representação cognitiva compartilhada universalmente por todos os seres humanos dotados de linguagem.	Realização audível de um conceito abstrato, variável de uma língua para outra; por exemplo, a noção de 'árvore' se manifesta como *arbor, dendron, Baum, 'mbira, tree* em latim, grego, alemão, tupi e inglês, respectivamente.
SINCRONIA	**DIACRONIA**
Estado do sistema linguístico num dado momento histórico; a sincronia é um construto teórico, feito com base nas deduções operadas pelo linguista.	Estudo da língua ao longo do tempo, observando seus diferentes estados, seus processos de mudança, documentados em textos escritos.
PARADIGMA	**SINTAGMA**
Repertório virtual de signos que o falante pode ativar para dele escolher os elementos na hora de se expressar.	Encadeamento concreto de signos na cadeia falada; reunião de signos decorrente dos processos mentais de seleção dentro do paradigma.

Ao traçar suas dicotomias, Saussure (ou o que nos chegou dele por meio de seus alunos) faz opções explícitas: a linguística que ele quer definir como uma ciência autônoma tem de se ocupar da langue, do sistema virtual, e não da parole, das falas individuais; tem de dar primazia absoluta à sincronia, o estado atual da língua, tal como os falantes de agora a conhecem e a empregam, sem que para descrever essa sincronia seja necessário recorrer ao estudo do passado da língua. Saussure, usando a metáfora do xadrez, muito recorrente em seu discurso, dizia que o importante é estudar o jogo tal como ele se apresenta num dado momento, as relações entre as peças tal como colocadas aqui e agora, e não procurar saber que sequência de lances permitiu que a partida assumisse o aspecto que tem aqui nesse instante — como o jogador conhece as regras do jogo, pode entender a situação atual do tabuleiro, do modo como ela se apresenta; da mesma forma, um falante da língua, conhecedor de suas regras, não precisa conhecer as origens do idioma para usá-lo. Recorrendo a uma metáfora mais contemporânea, poderíamos dizer que para Saussure a língua era como um programa de computador em sua versão mais nova: as pessoas podem muito bem aprender a usar esse programa, conhecer o funcionamento dele, sem precisar saber o que ele tem de novo ou diferente com relação às suas diversas versões passadas.

A linguística estrutural, inaugurada por Saussure, vai propor, então, um conceito de *língua* como uma entidade autônoma, um sistema fechado em si mesmo, que pode ser estudado sem se levar em conta qualquer interferência histórica, social, cultural etc. Como é fácil depreender pela tabela acima, as entidades da coluna da esquerda pertencem ao plano do *inteligível*, do *dedutível* a partir de abstrações, enquanto as da direita estão confinadas à caverna do sensorial, do audível, do articulável.

É graças a essa visão da "língua em si mesma e para si mesma" que o estruturalismo pôde elaborar conceitos como os de *fonema*, *morfema*, *sema* e *lexema*, que não se referem a nenhuma entidade física, a nenhum dado concreto da realidade, mas a unidades abstratas, componentes do *sistema* — igualmente abstrato e virtual — que seria uma língua.

Esse recurso à abstração é que torna tão difícil para os estudantes novatos de linguística a compreensão, por exemplo, do conceito de fonema. Tão difícil que muitos saem da faculdade diplomados sem serem capazes de explicar o conceito ou, pior, definindo fonema como "som da língua", equívoco teórico que se apresenta abundantemente nos livros didáticos disponíveis no mercado.

Embora o conceito de fonema esteja sendo abandonado por um número cada vez maior de estudiosos da fonética, convém entender o que está em jogo aqui para apreendermos o platonismo impregnado no estruturalismo. Se sairmos de gravador em punho Brasil afora e registrarmos o modo como as pessoas pronunciam a segunda consoante da palavra escrita MARÇO, vamos encontrar um leque de pronúncias diferentes: uma vibração simples com a ponta da língua; uma vibração múltipla com a ponta da língua ou com o fundo da garganta; uma aspiração forte, uvular; uma aspiração mais fraca, gutural; uma aspiração ainda mais fraca; uma retroflexão (o "R caipira"); uma quase vocalização (algo como "maiço"); e até mesmo uma realização zero (algo como "maço"), entre outras. Para cada uma dessas pronúncias a fonética científica criou um símbolo, que usamos (internacionalmente, aliás) para transcrever esses sons: [ɾ], [x], [ɹ], [ʀ], [ʁ], [h], [ɪ] etc. Esses, sim, são "sons da língua", mas não são *fonemas*! A ciência fonética vai dizer que eles são *fones* (ou, numa outra perspectiva, *alofones*).

E o que é um *fonema*, então? Aqui, passamos do nível do concreto, do gravável, do audível e do pronunciável — ou seja, da caverna de Platão — para o nível do abstrato, do apenas teorizável, para o mundo das formas puras. Os linguistas estruturalistas faziam a seguinte reflexão: todos esses diferentes *fones* (sons) não impedem que os falantes da língua reconheçam que se trata de uma mesma palavra, que se escreve oficialmente MARÇO e que significa, para todos eles, "o terceiro mês do ano". Com base nessa reflexão, os linguistas concluíam: "Então, existe no português brasileiro um *fonema* /ɾ/ que pode se materializar, concretamente, como os sons [ɾ], [x], [ɹ], [ʀ], [ʁ], [ɪ] etc." Observe que o fonema, entidade abstrata, vem representado entre barras oblíquas, enquanto os fones, sons de fato, são representados entre colchetes.

Na linguística estrutural, o sistema é composto de elementos que se *opõem*, que se *distinguem* uns dos outros. No nível fonológico (ou fonêmico), essa distinção será *semântica*, isto é, tomará como base o *significado* que os elementos em oposição têm para os falantes da língua (e não o *significante*). Assim, o fonema /s/ (presente em MARÇO) se distingue do fonema /ʃ/ que representamos na escrita com CH em MARCHO (do verbo *marchar*), por exemplo. Uma vez que todo falante da língua reconhece significados diferentes para *março* e *marcho*, podemos dizer que a oposição /s/ ≠ /ʃ/ se dá, aqui, no nível fonêmico. *Março* e *marcho* formam, então, o que se chama de *par mínimo*, isto é, uma dupla de palavras que se distinguem apenas por um único *fonema*. No entanto, a diferença entre a pronúncia ['baʃta] e a pronúncia ['basta] para o que se escreve BASTA não implica em mudança de significado — aqui, estamos no nível fonético e, portanto, [ʃ] e [s] não são fonemas, mas alofones do fonema /s/.

Para refinar (e complicar) um pouco mais: se um carioca pronuncia a palavra TIA como ['ʧia] e um pernambucano como ['tia] (sem "chiado"), essa diferença apenas no nível do *significante*, por não envolver mudança de *significado*, também não está no nível *fonológico*, mas apenas no nível *fonético* (no nível dos sons da língua, dos fones). Por isso, a fonologia estruturalista diz que não existe o fonema /ʧ/ em português (mas em inglês, sim, porque ele serve para distinguir palavras diferentes, como *cash* e *catch*; assim como em espanhol, com pares mínimos como *gato* e *gacho*).

Uma outra definição possível é a que identifica o *fonema* como uma *classe de sons*. Assim, retomando nosso exemplo, os sons [ɾ], [x], [ɹ], [ʀ], [ʁ] etc. formam uma classe de sons: o fonema /ɾ/.

Apesar de ter sido de grande valia para o entendimento do funcionamento da língua falada, a noção de *fonema* aos poucos foi se revelando problemática, justamente por causa de sua elevada dose de abstração, e hoje, como dissemos, já é descartada por muitos estudiosos da fonologia.

1.4 Platonismo gerativista: Chomsky

O linguista americano Noam Chomsky (n. 1928), que também é um destacado ativista político de combate ao imperialismo de seu país e à opressão que os Estados Unidos exercem sobre o resto do mundo, é o fundador de uma escola teórica extremamente influente conhecida como gerativismo ou gramática gerativo-transformacional. Declarando-se herdeiro do racionalismo de Descartes (a ponto de chamar sua teoria de *linguística cartesiana*), Chomsky se diferencia de Saussure num aspecto fundamental: enquanto o linguista suíço abstraía suas entidades virtuais de um mundo que ele chamava de *social* (mas, de fato, uma sociedade esquemática, abstrata), Chomsky se vincula à *psicologia cognitiva* e traça suas dicotomias separando o que se processa no cérebro do falante e o que se expressa concretamente na

realização falada. A cada tantos anos, Chomsky abandona alguns conceitos, propõe outros, constrói e reconstrói seu edifício teórico. Mesmo assim, o mecanismo epistemológico que ele emprega nessas (des)(re)construções é sempre o da dicotomia. Alguns de seus conceitos mais conhecidos são:

ESTRUTURA PROFUNDA	ESTRUTURA SUPERFICIAL
A estrutura profunda de uma expressão linguística é um construto teórico que tenta unificar diversas estruturas relacionadas. Por exemplo, as frases "Ernesto ama Célia" e "Célia é amada por Ernesto" significam, *grosso modo*, a mesma coisa e usam palavras semelhantes. Chomsky postulava que essas duas frases são *formas superficiais* diferentes geradas por uma *estrutura profunda* comum. Em versões posteriores de sua teoria, ele vai abandonar essa dicotomia inicial.	
COMPETÊNCIA	**DESEMPENHO**
A competência é o sistema de conhecimento linguístico que os falantes nativos de uma língua dominam. É o conhecimento que torna possível aos falantes produzir e entender um número infinito de frases em sua língua, capacitando-os a distinguir a *gramaticalidade* e a *agramaticalidade* de um enunciado em sua língua, isto é, de reconhecer se determinado enunciado está ou não bem formado segundo as regras da língua. Tal como a *langue* de Saussure, a competência de Chomsky é dedutível, é um construto teórico, e não temos acesso direto a ela. Quanto ao uso que os falantes fazem dessa competência, ele é definido por Chomsky como *desempenho* (*performance*). A linguística chomskiana é, portanto, uma tentativa de descrever (deduzir) essa competência a partir dos dados fornecidos pelo desempenho.	
PRINCÍPIOS	**PARÂMETROS**
Em versão mais recente da teoria chomskiana, todas as línguas humanas apresentam *princípios (universais)* que se manifestam em cada língua na forma de *parâmetros (particulares)*. Por exemplo, o princípio de *sujeito* é comum a todas as línguas, mas cada uma realiza esse princípio conforme *parâmetros* específicos: em algumas línguas, como o inglês, o sujeito se expressa por um pronome, obrigatoriamente; em outras, como o latim clássico, o sujeito vem expresso nas flexões verbais: amo, amas, amat, amamus, amatis, amant. Assim, quando uma língua deixa de expressar o sujeito somente pela flexão e passa a empregar obrigatoriamente os pronomes pessoais, diz-se que ela sofreu uma *mudança paramétrica*. Aqui a oposição de fundo é *universal* x *particular*.	
LÍNGUA-I	**LÍNGUA-E**
Distinção semelhante, mas não idêntica, à de competência e desempenho. As letras I e E se referem, respectivamente, a "interna"/"intensional" e "externa"/"extensional". A Língua-I é tomada como o objeto de estudo da linguística teórica; ela é o conhecimento linguístico mentalmente representado que um falante nativo da língua detém, sendo portanto um objeto mental. Com isso, a linguística se torna um ramo da psicologia. A Língua-E abarca todas as demais noções do que seja uma língua. A Língua-I, como objeto mental, é tudo o que interessaria a uma teoria linguística.	

A filiação de Chomsky à filosofia dualista platônica (e cartesiana) se evidencia em sua tentativa de resolver o que ele mesmo batizou de problema de Platão, isto é, a grande lacuna que existe entre *conhecimento* e *experiência* (mais uma dicotomia). A pergunta envolvida aqui é: como formamos o nosso conhecimento, que é tão poderoso, a partir de poucas experiências e dos poucos dados fornecidos pelo nosso ambiente? É o que o gerativismo chama de "axioma da pobreza de estímulo". Como uma criança pequena consegue, a partir de certa idade, produzir e compreender uma infinidade de construções linguísticas às quais ela jamais foi exposta?

No diálogo *Mênon*, de Platão, a resposta está na natureza *inata* do conhecimento, que deriva do fato de cada alma humana já trazer, de suas vidas passadas, lembranças do que aprendeu: é o que se chama de *anamnese*. De fato, para o platonismo não existe propriamente *conhecimento*, mas *re-conhecimento*, uma vez que as almas, antes de ocuparem corpos terrenos, habitaram o chamado Mundo das Ideias, onde todo o conhecimento é límpido, total e perfeito. Ao ocupar corpos humanos, materiais, as almas foram perdendo esse conhecimento total, que no entanto se manifesta quando elas são submetidas ao processo de anamnese, ou seja, de resgate da memória, como faz o mestre de Platão, Sócrates, no *Mênon*, ao mostrar que um menino iletrado conhecia os princípios básicos da geometria.

Para Chomsky, a resposta é semelhante, mas está no campo da biologia: a linguagem é uma faculdade inata, quer dizer, presente na nossa própria composição genética, de modo que nosso cérebro já vem programado para processar os poucos dados fornecidos pelo ambiente e, a partir deles, organizar em nossa mente toda a gramática de nossa língua. Para Chomsky, a metáfora do computador parece a mais adequada: já nascemos dotados de um *hardware* (que ele chama de gramática universal) que, para funcionar, só precisa de um *software*, de um programa operacional, fornecido pela língua falada no ambiente em que a criança nasce e é criada. Essa visão mecanicista da cognição humana tem sido duramente criticada nas últimas décadas.

A filiação de Chomsky à epistemologia dos filósofos gregos antigos é explicitada pelo próprio autor ao escrever (1975: 3-4):

> Uma das razões para estudar a linguagem — e para mim, pessoalmente, a mais premente delas — é o fato de ser tentador considerar a linguagem como um "espelho da mente", segundo a locução tradicional. Com isso não quero dizer simplesmente que os conceitos expressos e as distinções desenvolvidas na linguagem normal nos revelam os modelos de pensamento e de universo do "senso comum", construídos pela mente humana. Mais intrigante ainda, pelo menos para mim, é a possibilidade de podermos descobrir, ao estudar a linguagem, princípios abstratos que são universais por necessidade biológica e não por simples acidente histórico, e que derivam de características mentais da espécie.

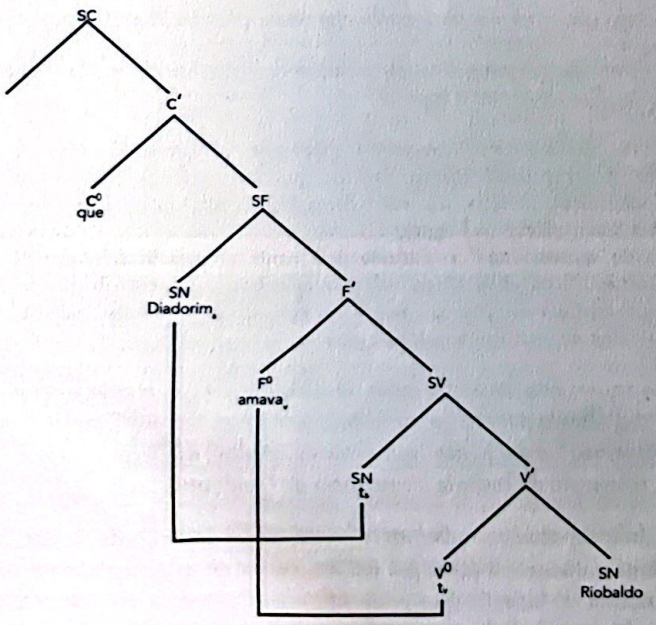
Encontramos aqui o esquema especular da filosofia antiga descrito acima: a linguagem como "espelho da mente" e, por isso, como espelho dos "modelos de pensamento e de universo"; o apelo aos "princípios abstratos" e à "necessidade biológica", junto com a relegação da história à categoria de "acidente".

Saussure e Chomsky são dois máximos representantes da escola de pensamento chamada de estruturalismo. A principal característica do estruturalismo é a desconsideração absoluta do sujeito, isto é, do falante como ser historicamente situado, membro de uma sociedade, participante de uma cultura, dotado de vontade política e capacidade de ação e decisão.

Os estruturalismos são também formalistas, na medida em que não se interessam pelo que é material, concreto, empiricamente coletável, etc., e concentram seus esforços na construção de um sistema abstrato, *formalizado* em regras. Saussure, por exemplo, traça uma diferença entre *forma* e *substância* que pode ser considerada como a ancestral longínqua da dicotomia *princípios e parâmetros* de Chomsky. Não por acaso, a linguística estruturalista vai eleger a *forma* como seu objeto de estudo e não a substância. E também não é por acaso que a palavra *forma* foi escolhida aqui, uma vez que, no *platonismo*, *forma* é "cada uma das realidades transcendentes que contêm a essência imaterial dos objetos concretos, realidades captáveis somente pelo intelecto que supera as impressões sensíveis (o mesmo que arquétipo ou ideia)" (*Houaiss*). O *arquétipo* platônico é o antepassado da *regra*, tal como entendida na gramática gerativa.

Para Chomsky, todo falante de uma língua é perfeito conhecedor de sua gramática, de modo que o linguista não precisa sair a campo coletando dados do uso: ele pode se servir de sua própria intuição linguística para deduzir as regras que põem sua língua em funcionamento.

A formalização de regras chega a seu extremo com o gerativismo chomskiano que, como já foi dito, concebe a linguagem humana como um código computacional, que pode ser descrito por meio de regras muito semelhantes às expressões matemáticas ou, mais precisamente, a sistemas de rede computacionais (não por acaso, essas formalizações são classificadas como pertencentes a uma linguística computacional):



Subida do sujeito para a posição de especificador de SC (sintagma complementizador) (reproduzido de Berlinck, Augusto e Scher, in Mussalim e Bentes, 2001: 218).

Uma das principais críticas dirigidas à gramática gerativa é seu apego exclusivo à sintaxe, às frases isoladas, como se elas fossem suficientes para explicar toda a complexidade da linguagem humana. E, de forma ainda mais reducionista, a uma sintaxe que independe de todo processo de interação social, que desconsidera a natureza primordialmente discursiva e dialógica da linguagem, uma faculdade humana que existe para conferir sentido à experiência e, com base nela, produzir conhecimento e interagir com os demais seres humanos. Não surpreende, portanto, que Chomsky trabalhe no MIT (Massachusetts Institute of Technology), o mais importante centro mundial de pesquisa tecnológica: a *língua* com a qual ele trabalha é um objeto altamente esquemático e regular, sem comprovação empírica na realidade dos usos — tanto quanto a matemática, a linguística chomskiana busca um rigor axiomático que não se aplica à linguagem humana natural.

Podemos fazer à linguística chomskiana a mesma pergunta que foi feita a Albert Einstein sobre a matemática: "Como é que a matemática, sendo afinal de contas um produto do pensamento humano, independente da experiência, é tão admiravelmente apropriada aos objetos da realidade?" E oferecer a essa pergunta a mesma resposta dada pelo mesmo Einstein: "Na medida em que se referem à realidade, as leis da matemática [*e da linguística chomskiana*] não admitem certeza; e na medida em que admitem certeza, elas não se referem à realidade".

A recusa da linguística gerativa em considerar a experiência do falante como ser social e historicamente situado é explicitada pelo próprio Chomsky numa de suas citações mais famosas:

> A teoria linguística se interessa primordialmente por um falante-ouvinte ideal, numa comunidade de fala completamente homogênea, que conhece sua língua perfeitamente e não é afetado por condições gramaticalmente irrelevantes tais como limitações de memória, distrações, desvios de atenção e interesse, e erros (aleatórios ou característicos) ao aplicar seu conhecimento da língua no desempenho real (Chomsky, 1965: 3-5).

Com isso, a teoria chomskiana adquire um ar "místico", na medida em que — assim como na filosofia platônica — a *língua*, tal como o gerativismo a entende, não pertence ao mundo do sensível, do audível, do legível, mas a uma esfera transcendente, a um Mundo das Ideias (e é de *ideia* que vem *ideal* em "falante ideal") habitado pelas Formas perfeitas do platonismo. Há autores que vêm denunciando alguns dos postulados centrais do gerativismo — sobretudo a gramática universal e o inatismo da linguagem — como mitos. Christiansen e Chater (2008: 1), por exemplo, escrevem:

> Supõe-se amplamente que o aprendizado humano e a estrutura das línguas humanas sejam intimamente relacionados. Sugere-se com frequência que essa relação derive de uma específica dotação biológica para a linguagem, a qual codifica princípios da estrutura linguística universais, mas comunicativamente arbitrários (uma gramática universal ou GU). Como essa GU poderia ter evoluído? Afirmamos que a GU não poderia ter emergido nem por adaptação biológica nem por processos genéticos não adaptativos, resultando num *problema lógico da*

evolução da linguagem. Especificamente, na medida em que os processos de mudança linguística são muito mais rápidos do que os processos de mudança genética, a linguagem constitui um "alvo móvel" tanto ao longo do tempo quanto entre as diferentes populações humanas e, por conseguinte, ela não tem como oferecer um ambiente estável ao qual os genes da linguagem poderiam ter se adaptado. Concluímos que uma GU biologicamente determinada não é viável do ponto de vista evolutivo. Em vez disso, a motivação original para a GU — a mescla entre aprendizes e línguas — emerge porque a linguagem tem sido moldada para se adequar ao cérebro humano, e não vice-versa. Acompanhando Darwin, consideramos a linguagem mesma como um "organismo" complexo e interdependente, que evolui sob pressões seletivas a partir do aprendizado humano e dos mecanismos de processamento. Ou seja, as línguas mesmas são moldadas por severa pressão seletiva da parte de cada geração de usuários e aprendizes da língua. Isso sugere que aspectos aparentemente arbitrários da estrutura linguística podem resultar de vieses gerais no aprendizado e no processamento, derivando da estrutura dos processos de pensamento, de fatores perceptomotores, de limitações cognitivas e da pragmática.

Por sua vez, Evans e Levinson (2009: 429) afirmam:

As línguas são muito mais diversas em estrutura do que os cientistas cognitivos em geral reconhecem. Uma suposição amplamente difundida entre os cientistas cognitivos, derivada da tradição gerativista na linguística, é a de que todas as línguas são parecidas com o inglês, mas com sistemas sonoros e vocabulários diferentes. O quadro verdadeiro é muito diferente: as línguas diferem tão fundamentalmente uma da outra em cada nível de descrição (som, gramática, léxico, significado) que é dificílimo encontrar uma única propriedade estrutural sequer que elas compartilhem. As defesas da gramática universal, argumentamos aqui, são ou empiricamente falsas, não falsificáveis ou enganadoras ao se referirem a tendências mais do que a universais estritos. As diferenças estruturais deveriam, ao contrário, ser aceitas pelo que são e integradas numa nova abordagem da linguagem e da cognição que coloque a diversidade no centro do palco.

É muito grave essa concepção errônea de que as diferenças entre as línguas são meramente superficiais e de que podem ser resolvidas postulando-se um nível formal mais abstrato no qual desaparecem as diferenças entre as línguas individuais: ela hoje permeia muito do trabalho que se faz em psicolinguística, em teorias da evolução linguística, aquisição de linguagem, neurocognição, análise sintática e reconhecimento de fala, e praticamente em todos os ramos das ciências cognitivas (p. 429).

Quanto à ideia de universais linguísticos, Michael Tomasello (2003: 18-19) afirma:

É claro que existem universais linguísticos. Só que não são universais de forma — isto é, não são tipos particulares de símbolos linguísticos, categorias gramaticais ou construções sintáticas — porém, bem mais, são universais de comunicação, cognição e fisiologia humana. Uma vez que todas as línguas são usadas por seres humanos com vidas sociais semelhantes, todas as pessoas têm a necessidade de resolver em suas línguas certos tipos de tarefas comunicativas, tais como referir-se a entidades específicas ou predicar coisas sobre essas entidades. Todos os seres humanos têm também as mesmas ferramentas básicas para realizar tais tarefas — símbolos linguísticos, marcadores sobre esses símbolos, ordenação de símbolos, padrões prosódicos, e certos processos recorrentes de gramaticalização. É isso que leva a alguns universais linguísticos: dispor, por exemplo, de nomes e verbos para a expressão da referência e da predicação. Tais universais são fenômenos emergentes, fundados, em última instância, nos universais da cognição humana, das necessidades comunicativas humanas, do processamento vocoauditório humano [...]

Outro problema central da gramática gerativa é o fato de ter sido concebida por um falante nativo do *inglês americano urbano de prestígio*, de modo que seus postulados e suas técnicas de análise decorrem em grande medida desse fato — social e histórico. Com toda probabilidade, se Chomsky fosse falante de basco, malaio ou japonês, sua teoria apresentaria uma aparência totalmente diversa. Por isso, muitas das classificações de *agramaticalidade* (inaceitabilidade) aplicadas a determinadas construções do inglês decorrem do fato, banal, dessas construções não pertencerem à variedade falada pelo linguista. Por isso, Evans e Levinson não hesitam em denunciar o etnocentrismo dessas concepções:

Como surgiu essa difundida concepção errônea de uniformidade linguística? Em parte, ela pode ser atribuída simplesmente ao etnocentrismo — a maioria dos cientistas cognitivos, linguistas incluídos, falam somente as línguas europeias mais conhecidas, todas elas aparentadas na estrutura (p. 430).

Podemos dizer então do gerativismo o que É. Benveniste disse acerca da lógica aristotélica: é toda baseada nas *estruturas gramaticais da língua grega* — se Aristóteles falasse outra língua, com regras gramaticais bem diferentes das do grego, os postulados de sua lógica seriam também inteiramente outros. Isso deixa claro que, desde seu aparecimento, a *lógica*, como disciplina, é que esteve sujeita ao funcionamento da língua, e não o contrário — embora tanta gente teime em analisar a língua pela ótica da lógica clássica.

A essa preponderância do *formal* vão se opor diversos pensadores ao longo do século XX, que se esforçarão por incluir na linguística o uso real da língua por parte dos falantes e a perspectiva teórica de que a linguagem tem uma função (bem como um funcionamento) eminentemente sociocultural e sociocognitiva, coerente com a própria natureza social e gregária do ser humano, que Aristóteles já tinha definido como *zôon políticon*, "animal político".

Alguns fatores históricos explicam o sucesso da linguística estruturalista na primeira metade do século XX. Podemos começar mencionando um forte sentimento *antigermânico* que imperava na Europa no final do século XIX e início do XX, decorrente das pretensões expansionistas da Alemanha e que culminaram na eclosão da I Guerra Mundial (1914-1918). Por isso é que, ao término da II Guerra Mundial (1945), Saussure começou a ser considerado como o "pai da linguística moderna", um título que até hoje os linguistas alemães se recusam a atribuir ao pensador suíço, uma vez que é perfeitamente possível vincular as origens da linguística moderna ao trabalho dos comparatistas e dos neogramáticos, fundadores da linguística histórica, cujos nomes mais importantes são todos alemães: August von Schlegel (1767-1845), Friedrich Diez (1794-1876), os irmãos Jacob (1785-1863) e Wilhelm Grimm (1786-1859), Franz Bopp (1791-1867), August Schleicher (1821-1868), August Leskien (1840-1916), Hugo Schuchardt (18421927), Berthold Delbrück (1862-1922), Hermann Paul (1846-1921), Hermann Osthoff (1847-1909), Karl Brugmann (1849-1919), Wilhelm Meyer-Lübke (1861-1936), entre tantos outros.

Muitos desses filólogos e linguistas são, reconhecidamente ou não, devedores dos trabalhos pioneiros do dinamarquês Rasmus Rask (1787-1832), talvez o verdadeiro merecedor do título de fundador da ciência linguística moderna, com suas intuições brilhantes e suas pesquisas de campo, absolutamente inovadoras para a época, em que os filólogos se limitavam a pesquisas de gabinete, debruçados sobre os manuscritos antigos.

Rask empreendeu expedições científicas para investigar as línguas faladas na Suécia, Finlândia, Rússia, Tartária, Pérsia, Índia, Ceilão... Escreveu seus trabalhos em dinamarquês, sueco, alemão e inglês. Publicou gramáticas de línguas como o espanhol, o anglo-saxão, o frísio, o zende, o italiano, o sami. Foi o primeiro a postular um parentesco entre as diferentes línguas da Europa. Perto do fim da vida, dominava mais de vinte e cinco línguas e dialetos, tendo se dedicado ao estudo aprofundado do dobro disso. Foi o precursor das hipóteses que mais tarde receberiam o nome de "leis de Grimm", sobre a mudança fonética, tendo comparado apenas o teutônico e o grego. Desse modo, foi o pioneiro dos trabalhos que, logo adiante, levariam ao estabelecimento definitivo da família indo-europeia.

O grupo de linguistas denominados neogramáticos (tradução inadequada do alemão *Junggramatiker*, "jovens gramáticos") se formou e atuou na cidade alemã de Leipzig. Ora, foi precisamente em Leipzig que Saussure estudou e onde obteve seu grau de doutor. Também estudou em Berlim, antes de retornar a sua Genebra natal para dar seus famosos cursos. Sua formação, portanto, se fez toda na Alemanha, nas escolas de pensamento linguístico que predominavam naquele país. Sua suposta "rebeldia" contra a linguística histórica e a favor de uma linguística obsessivamente sincrônica não esconde, no fundo, sua dependência e sua dívida para com seus mestres alemães[1]. Tanto quanto o estruturalismo saussuriano define a língua como um objeto autônomo, independente dos indivíduos que a falam, também os neogramáticos concebiam, por exemplo, o sistema fonológico das línguas como sujeitos a "leis" que eram seguidas "cegamente" por esse mesmo sistema, considerado totalmente desvinculado dos demais níveis (morfológico e sintático) da língua e, portanto, desvinculado dos seres vivos, históricos e sociais que falam essa língua. Nas palavras de Osthoff e Brugmann (1878: xiii):

> Toda mudança fonética, *na medida em que se realiza mecanicamente e por si mesma*, ocorre segundo *leis sem exceção*, isto é, a evolução dos sons é sempre a mesma em todos os membros de uma comunidade linguística [...] [grifos meus].

[1] É sempre necessário recordar que o livro atribuído a Saussure, *Cours de linguistique générale* (*Curso de linguística geral*), foi publicado em 1916, três anos após a morte do linguista, com base em anotações de seus discípulos Bally e Sechehaye. Assim, o que temos é, de fato, uma espécie de "evangelho" saussuriano, que não foi escrito por ele, mas por seus seguidores. Por isso, quando se fala de "ideias saussurianas", "teorias saussurianas", "conceitos saussurianos" etc., o que está em jogo é o discurso presente no *Curso* que não pode ser vinculado estrita e diretamente ao mestre de Genebra. Além disso, foi publicado em 2002 um livro intitulado *Écrits de linguistique générale* (*Escritos de linguística geral*) com manuscritos inéditos de Saussure, com alguns trechos que põem em dúvida certas asserções categóricas presentes no *Curso*.

Outro fator histórico importante para o florescimento e desenvolvimento do estruturalismo na linguística ocidental foi a Revolução Russa de 1917 e, mais adiante, a ditadura stalinista, que fecharam a União Soviética em si mesma e impediram que as teorias científicas produzidas por lá fossem divulgadas em outros países. Se a linguística moderna tivesse entrado em contato, já nos anos 1920, com a filosofia da linguagem e a psicologia produzidas na União Soviética, certamente os rumos dos estudos sobre as línguas teriam sido bem diferentes. Postulados que só surgiriam no Ocidente a partir dos anos 1950 e, como no caso da sociolinguística, nos anos 1960, já tinham sido lançados por pensadores russos algumas décadas antes. É o que vamos ver a seguir.

1.5 Volóshinov: contra o dualismo na linguagem

Valentin Nikolaévitch Volóshinov [Валентин Николаевич Волошинов] nasceu em 1895 e morreu (de tuberculose) em 1936. Pertenceu ao famoso círculo de intelectuais que se reunia em torno do filósofo Mikhail Bakhtin (1895-1975). Muitos dos membros do círculo sofreram com a perseguição promovida pelo stalinismo a partir dos anos 1930. Assim, por exemplo, o crítico literário Pavel N. Medvedev (1892-1938) foi preso e "desapareceu" para sempre. O próprio Bakhtin foi condenado a trabalhos forçados na Sibéria. No entanto, como tinha a saúde já muito debilitada, sua pena foi comutada em exílio interno numa cidade obscura do interior. Durante algumas décadas, obras de autoria de Volóshinov e de Medvedev foram atribuídas a Bakhtin. Mais recentemente, no entanto, os estudiosos têm preferido devolver a autoria dessas obras a seus verdadeiros produtores. É essa posição que vamos assumir aqui.

Volóshinov é o autor de pelo menos dois livros muito importantes e influentes: O *freudismo* (1927) e *Marxismo e filosofia da linguagem* (1929). Ambos têm publicação brasileira sob o nome de Bakhtin, embora na capa do segundo também apareça o nome de Volóshinov entre parênteses. As duas obras foram publicadas em russo no final dos anos 1920, mas só na década de 1960 tiveram suas primeiras traduções em línguas ocidentais, com autoria sempre atribuída a M. Bakhtin.

Uma das principais contribuições de Volóshinov ao pensamento linguístico contemporâneo é sua crítica às duas grandes concepções de língua e de linguagem que, segundo ele, sempre dominaram os estudos filológicos, gramaticais e linguísticos até sua época.

A primeira dessas concepções de língua é chamada por Volóshinov de subjetivismo idealista. É a percepção da língua como uma "atividade mental", em que o psiquis-

mo individual constitui a fonte da língua. Volóshinov sintetiza essa concepção nas seguintes afirmações ([1929] 1986: 48):

1. A língua é uma atividade, um processo criativo ininterrupto de construção ("*energeia*"), que se materializa sob a forma de atos de fala individuais.
2. As leis de criação linguística são essencialmente leis individual-psicológicas.
3. A criação linguística é uma criação racional análoga à criação artística.
4. A língua, na qualidade de produto acabado ("*ergon*"), na qualidade de sistema estável (léxico, gramática, fonética) se apresenta como um depósito inerte, tal como a lava esfriada da criação linguística, abstratamente construída pelos linguistas em vista de sua aquisição prática como ferramenta pronta para o uso.

Volóshinov identifica o pensador alemão Wilhelm von Humboldt (1767-1835) como um importante (talvez o maior) representante e defensor dessa primeira concepção de língua. No entanto, é também possível incluir Noam Chomsky entre os que veem a língua na perspectiva de um "subjetivismo idealista". De fato, conforme discutimos acima, Chomsky propõe classificar a linguística como um ramo da psicologia cognitiva, baseia suas análises na produção verbal de um "falante ideal", abstraído de toda realidade histórica e social, e empreende a busca de uma "gramática universal", igualmente imune às investidas da ideologia e da vida social dos seres humanos. Publicado em 1929, um ano depois do nascimento de Chomsky, o livro de Volóshinov parecia antever o sucesso da linguística cartesiano-platônica do pensador americano recém-nascido e já lhe oferecia uma crítica acabada. Na verdade, uma vez que Chomsky pertence a uma tradição que remonta a Descartes e a Platão, a crítica de Volóshinov nada tinha de profética, mas simplesmente de demonstração das inadequações dessa antiga linhagem teórica como explicação dos fenômenos de linguagem. Mas, ainda assim, é surpreendente ler palavras como as seguintes, que parecem dirigidas explicitamente ao fundador da gramática gerativa: ao sintetizar os postulados do *subjetivismo idealista*, Volóshinov escreve (p. 49) que, na concepção dessa tendência de análise da linguagem,

> o psiquismo individual constitui a fonte da língua. As leis da criação linguística [...] são essencialmente leis individual-psíquicas e elas é que devem ser estudadas pelo linguista e pelo filósofo da linguagem. Esclarecer o fenômeno linguístico significa reconduzi-lo a um ato de criação individual refletido (até mesmo racional). Todo o resto da tarefa do linguista tem um caráter apenas preliminar, construtivo, descritivo, classificador; consiste simplesmente em preparar a explicação exaustiva do fato linguístico como proveniente de um ato de criação individual, ou então em servir os objetivos práticos da aquisição de uma língua acabada.

A referência ao psíquico, ao racional e à aquisição nos remete imediatamente a postulados centrais da linguística chomskiana. Volóshinov também menciona explicitamente, em sua crítica, o cartesianismo (p. 87). A comparação que fizemos acima entre a noção de língua do gerativismo e a matemática também é abordada por Volóshinov, que faz inclusive uma menção à ideia de gramática universal:

> A ideia de uma língua *convencional, arbitrária*, é característica de toda a corrente racionalista, assim como o paralelo estabelecido entre o código *linguístico* e o código *matemático*. Não é a relação do signo com a realidade que ele reflete ou com o indivíduo que o gera, mas a relação *de um signo com* outro signo no interior de um sistema *fechado*, e entretanto aceito e integrado, que interessa ao espírito dos racionalistas, orientado para a matemática. Em outros termos, só lhes interessa a *lógica interna* do próprio sistema de signos; este é considerado como uma álgebra, totalmente independente das significações ideológicas que se prendem a ele. [...] o signo matemático era, para os racionalistas, o signo por excelência, o modelo semiótico, inclusive para a língua. É exatamente isso que encontramos claramente expresso na ideia leibniziana da gramática universal (p. 88).

A segunda concepção de língua criticada por Volóshinov é a que ele chama de objetivismo abstrato. É basicamente a concepção da língua como um *sistema* de *formas* passíveis de descrição. A crítica se dirige agora explicitamente a Saussure e ao estruturalismo que então nascia a partir dos postulados saussurianos. Segundo o pensador russo, a linguística saussuriana, que acredita distinguir-se dos procedimentos da filologia tradicional, na verdade só faz reiterá-los e perpetuá-los, ao desdenhar a produção individual dos falantes (a *parole*) e ao se concentrar num construto teórico abstrato, homogêneo, impossível de verificação empírica (a *langue*, ou sistema, ou sincronia). Tal como faz para a primeira concepção tradicional, Volóshinov também sintetiza o "objetivismo abstrato" numa série de postulados (p. 57):

1. A língua é um sistema estável, imutável, de formas linguísticas submetidas a uma norma fornecida tal e qual à consciência individual e peremptória para esta.
2. As leis da língua são essencialmente leis linguísticas específicas que estabelecem vínculos entre os signos linguísticos no interior de um sistema fechado. Essas leis são objetivas em relação a *toda* consciência subjetiva.
3. Os vínculos linguísticos específicos nada têm que ver com valores ideológicos (artísticos, cognitivos ou outros). Não se encontra na base dos fatos de língua nenhum motor ideológico. Entre a palavra e seu sentido não existe vínculo natural e compreensível para a consciência, nem vínculo artístico.
4. Os atos de fala individuais constituem, do ponto de vista da língua, simples refrações ou variações fortuitas ou mesmo deformações das formas normalizadas. Mas são justamente esses atos de fala individuais que explicam a mudança histórica das formas da língua; enquanto tal, a mudança é, do ponto de vista do sistema, irracional e mesmo desprovida de sentido. *Entre o sistema da língua e sua história não existe nem vínculo nem comunhão de motores. O sistema e sua história são estranhos um à outra.*

Volóshinov vai opor a essas duas concepções a urgência de se considerar a língua como uma atividade social, em que o importante não é o *enunciado*, o *produto*, mas sim a *enunciação*, o *processo* verbal. Para Volóshinov, a língua é — tal como para Saussure — um fato social, cuja existência se funda nas necessidades da comunicação. Mas, contrariamente à linguística saussuriana e pós-saussuriana, que faz da língua um objeto abstrato ideal (um "arco-íris imóvel sobre o fluxo da língua", como escreve Volóshinov), que se consagra à língua como sistema sincrônico homogêneo e rejeita suas manifestações (a fala, *parole*) como individuais, Volóshinov enfatiza precisamente a fala, a *parole*, a *enunciação*, e afirma sua natureza social, não individual: a *parole* está indissoluvelmente ligada às condições de comunicação, que estão sempre ligadas às estruturas sociais.

Na análise essencialmente marxista de Volóshinov, todo signo é ideológico. A ideologia é um produto das estruturas sociais. Portanto, toda modificação da ideologia acarreta uma modificação da língua. A evolução da língua obedece uma dinâmica conotada positivamente, ao contrário da concepção saussuriana. A variação é inerente à língua e reflete variações sociais (e nessas afirmações Volóshinov se antecipou em meio século à sociolinguística, sem cair no dualismo que vai caracterizar a teoria da variação e mudança de Labov, como veremos adiante). Se é verdade que a mudança obedece, em parte, leis internas da língua, o fato é que essa mudança é regida *sobretudo* por leis externas, de natureza *social*. O signo dialético, movente, vivo, se opõe ao "sinal" inerte que se depreende da análise da língua como sistema sincrônico abstrato.

A palavra-chave da linguística voloshinoviana é diálogo. Só existe língua onde houver possibilidade de interação social, dialogal. A língua não reside na mente do falante, nem é um sistema abstrato que paira acima das condições sociais. A língua é um trabalho empreendido conjuntamente pelos falantes, é uma atividade social, é enunciação. A enunciação, compreendida como uma réplica do diálogo social, é a unidade de base da língua, tanto do discurso interior (diálogo consigo mesmo) quanto do discurso exterior. A natureza da língua é essencialmente dialógica, e isso se reflete nas próprias estruturas linguísticas: a sintaxe é narrativa, toda predicação é um relato, é algo que alguém conta a alguém. A enunciação é de natureza social, portanto ideológica. Ela não existe fora de um contexto social, já que todo falante tem um "horizonte social". Temos sempre um interlocutor, ainda que seja potencial. O falante pensa e se expressa para um auditório social bem definido. Se a língua é determinada pela ideologia, a consciência (portanto, o pensamento), a "atividade mental", que são condicionadas pela linguagem, são modeladas pela ideologia. A mente é um produto social — e nisso Volóshinov se aproxima de um compatriota e contemporâneo seu, o psicólogo Lev Vygotsky (1896-1934), que postulou "a construção social da mente". Tanto Volóshinov quanto Vygotsky parecem tributários do pensamento marxista de Georgi V. Plekhánov (1856-1918) autor precisamente, entre outras obras, de um ensaio sobre *O desenvolvimento da visão monista da história* (1895)[2].

Já em *O freudismo*, em que critica duramente a recém-nascida psicanálise freudiana, Volóshinov enuncia de forma muito clara o eixo teórico fundamental de sua concepção de linguagem:

> Nenhuma enunciação verbalizada pode ser atribuída exclusivamente a quem a enunciou: é *produto da interação entre falantes* e, em termos mais amplos, produto de toda uma situação social em que ela surgiu. [...] Todo produto da linguagem do homem, da simples enunciação vital a uma complexa obra literária, em todos os momentos essenciais, é determinado não pela vivência subjetiva do falante, mas pela situação social em que soa essa enunciação. [...] O que caracteriza precisamente uma dada enunciação — a escolha de certas palavras, certa teoria da frase, determinada entonação da enunciação — é a expressão da relação recíproca entre os falantes e todo o complexo ambiente social em que se desenvolve a conversa. As próprias "vivências psíquicas" do falante, cuja expressão tendemos a ver nessa enunciação, são de fato apenas uma interpretação unilateral, simplificada e cientificamente incorreta de um fenômeno social mais complexo. É uma espécie de "projeção" através da qual investimos (projetamos) na "alma individual" um complexo conjunto de inter-relações sociais. A palavra é uma espécie de "cenário" daquele convívio mais íntimo em cujo processo ela nasceu, e esse convívio, por sua vez, é um momento do convívio mais amplo do grupo social a que pertence o falante. Para compreender esse cenário, é indispensável restabelecer todas aquelas complexas inter-relações sociais das quais uma dada enunciação é a interpretação ideológica. A questão não muda se em vez de discurso exterior temos discurso interior. Esse discurso também pressupõe o ouvinte eventual, constrói-se voltado para ele. O discurso interior é tanto um produto e expressão do convívio social quanto o discurso exterior[3].

A refutação do *dualismo* na abordagem da língua, dualismo que, conforme vimos, é uma herança da filosofia platônica, é explicitada por Volóshinov ao opor a ele o monismo materialista. Segundo o autor, somente o monismo materialista é capaz de dar uma solução dialética para superar os problemas trazidos pela separação entre o psíquico e o ideológico:

> Em todo ato de fala, a atividade mental subjetiva se dissolve no fato objetivo da enunciação que tomou forma, enquanto a fala enunciada se subjetiviza no ato de decodificação que deve, mais cedo ou mais tarde, provocar a codificação de uma réplica. Cada palavra, como sabemos, se apresenta como uma arena em miniatura onde se entrecruzam e lutam os acentos sociais de orientação contraditória. A palavra se revela, na boca do indivíduo, como o produto da interação viva das forças sociais.
> É assim que o psiquismo e a ideologia se impregnam mutuamente no processo único e objetivo das relações sociais. ([1929] 1986: 40-41)

1.6 O RESGATE DO DISCURSO

As ideias de Volóshinov (indevidamente atribuídas a Bakhtin), bem como as de Vygotsky, só se tornarão conhecidas fora da Rússia no final dos anos 1970, quando vão ganhar suas primeiras traduções em línguas ocidentais. A partir daí, seu impacto será enorme e suscitará um amplo movimento de reflexão crítica dos postulados centrais do estruturalismo clássico saussuriano e do neoestruturalismo chomskiano.

Uma característica central das escolas de pensamento linguístico que vão se opor aos estruturalismos é o resgate da *parole* — a fala concreta, a língua em uso — do porão

[2] O texto completo dessa obra, em tradução para o inglês, está disponível em <http://www.marxists.org/archive/plekhanov/1895/monist/index.htm>. Agradeço aqui a Roxane Rojo pela sugestão da possível influência de Plekhánov sobre Volóshinov e Vygotsky.

[3] Citamos a tradução brasileira (diretamente do russo, por Paulo Bezerra): *O freudismo*, São Paulo: Martins Fontes, 2001, pp. 79-80.

escuro (ou caverna?) aonde tinha sido lançada pelo estruturalismo clássico, interessado somente na *langue*, no sistema abstrato, metafísico, construído pelo linguista sob a forma de um sistema computacional, algébrico. Sob a designação de discurso — palavra dotada de uma ampla gama de sentidos —, essa língua em uso se tornará o objeto primordial de diversas disciplinas e transdisciplinas e de várias escolas de investigação do fenômeno linguístico como fenômeno social:

- análise da conversação,
- análise do discurso,
- antropologia linguística,
- etnografia da comunicação,
- linguística sociocognitiva,
- linguística textual,
- pragmática linguística,
- semiótica,
- sociolinguística interacional,
- sociolinguística variacionista,
- sociologia da linguagem,
- sociointeracionismo,
- além de todas as diferentes teorias designadas pelo termo geral de funcionalismo.

A própria designação de muitas dessas linhas de investigação mostra seu caráter essencialmente transdisciplinar, ao contrário da linguística estrutural, que se esforçou por ser uma *ciência* autônoma, dentro do espírito positivista que caracterizou o pensamento ocidental no final do século XIX e início do XX. O estruturalismo transformou a língua numa "coisa em si" — objetivo explicitado pelas últimas palavras do *Curso de linguística geral*, livro em que se tentou compilar os ensinamentos de Saussure: "*A linguística tem por único e verdadeiro objeto a língua considerada em si mesma e por si mesma*". Para tanto, foi preciso desvincular o estudo da língua da *história* da língua e, por conseguinte, da história de seus falantes — daí a rejeição saussuriana da *diacronia*, o passado da língua, em favor da *sincronia*, um construto abstrato. Renegando o falante e sua inserção numa comunidade, a linguística estrutural renegou no mesmo gesto o social, o cultural, o político, o ideológico — em suma, tudo o que é caracteristicamente *humano*. Somente assim foi possível dar à linguística um caráter *científico*, um *status* de ciência plena, como se a linguagem humana e as línguas particulares fossem objetos semelhantes aos estudados pelas ciências naturais como a física, a química, a geologia etc. Apesar dos grandes avanços conquistados por essa linguística "científica" no entendimento das línguas, ela acabou topando com diversos obstáculos e muitos impasses, que só puderam ser transpostos com a reintrodução do *humano* e, portanto, das ciências humanas, na investigação do fenômeno linguístico.

Ao definir a *língua* como um sistema abstrato ou como o desempenho de um falante ideal, a linguística estruturalista (saussuriana e chomskiana) cria o que o sociólogo francês

Pierre Bourdieu chamou de "ilusão de comunismo linguístico". Nessas teorias, a língua, sendo atribuída a um "social" indefinido e abstrato (Saussure) ou a um "falante-ouvinte ideal" (Chomsky), deixa de ser o que ela de fato é na vida social: uma "arena em miniatura", como escreveu Volóshinov, palco de conflitos e disputas entre os membros da sociedade, marcado por negociações de interesses, valores e poderes (reais e simbólicos).

Outro problema detectado nessa "língua" é o fato de, ao fim e ao cabo, ela ser muito semelhante ao modelo idealizado de língua que figura na gramática tradicional. Ao tentar se elevar acima das normas sociais, o sistema que se constrói no empreendimento estruturalista acaba por se parecer muito com a norma de prestígio que vigora em determinada sociedade. Como já mencionamos, muitos rótulos de "agramaticalidade" atribuídos pelos gerativistas a certas construções sintáticas decorrem exclusivamente do fato de tais construções não fazerem parte da variedade linguística do pesquisador. E não poderia ser de outro modo: entranhada na vida social e cultural, a língua — mesmo quando forçada a se tornar um objeto ideal, uma "forma pura" — tem obrigatoriamente de assumir alguma feição, tem de se parecer com alguma das variedades reais que circulam no meio social. Uma vez que o linguista é membro das classes sociais mais letradas, seu modelo de língua, por mais abstrato, vai trazer as marcas da sua variedade sociolinguística. É isso o que leva Bourdieu a afirmar:

> A língua segundo Saussure, código ao mesmo tempo legislativo e comunicativo que existe e subsiste independentemente de seus usuários e de suas utilizações, possui de fato todas as propriedades comumente atribuídas à língua oficial (1996: 31).

Tentando romper com a tradição normativa, que sempre definiu seu objeto de estudo e de prescrição com base na escrita literária consagrada, o estruturalismo preconizou o estudo da língua falada. No entanto, dado o caráter ideológico do processo de construção da *língua/sistema/competência*, a linguística estrutural não teve como escapar do influxo milenar da escrita literária sobre os estudos da linguagem. Assim, conforme analisa James Milroy:

> Nos primórdios da gramática transformacional, era bastante comum ver sentenças que eram gramaticais em variedades regionais e em registros coloquiais serem marcadas com um asterisco como — simplesmente — agramaticais em inglês, sem nenhuma explicação do motivo por que eram agramaticais. Alguns desses linguistas pareciam ter um senso de observação bastante limitado: pareciam não estar conscientes das formas estruturais de nenhuma variedade que não o padrão literário. Assim, por exemplo, sentenças do tipo *the eggs is cracked* ("os ovos está quebrado"), com SN plural seguido de um verbo que seria singular no padrão, eram às vezes marcadas como agramaticais, muito embora tais sentenças sejam regulares (isto é, governadas por regras) em diversas variedades regionais. De igual modo, sequências como *they've done it last year* ou *he did it already*[4] foram consideradas por linguistas britânicos como

[4] O autor se refere aqui às regras da gramática normativa do inglês que prescrevem usos específicos para os dois tempos do passado, o simples (*I did*) e o composto (*I have done*). Os exemplos dados contrariam essas prescrições normativas.

agramaticais, muito embora elas sem dúvida ocorram regularmente tanto no inglês britânico quanto no americano. Acho que essa tendência dos teóricos de confiar numa única variedade (codificada) já é suficientemente conhecida e não preciso dar exemplos numerosos para apoiar meu argumento aqui. Note-se, porém, o que tudo isso implica: sequências que são gramaticais nos estilos formais do inglês padrão, literário, são tidas como "gramaticais", enquanto sequências de ocorrência regular em outras variedades são gramaticais somente na medida em que coincidem com o padrão (formal, literário); do contrário, estão fadadas a serem "agramaticais". É difícil ver qualquer diferença *em princípio* entre essas opiniões de gramaticalidade e as do público geral, e elas parecem ser igualmente dependentes da ideologia do padrão. Por algumas décadas, suposições sobre o *input* legítimo para a análise linguística implicitamente sustentaram o *status* superior e a alta saliência do inglês padrão e podem até ter contribuído de certa forma para a *manutenção* dessa variedade (2001: 545).

1.7 Breve crítica à sociolinguística variacionista

Os trabalhos de William Labov (n. 1927) iniciados na década de 1960, fundadores da chamada sociolinguística variacionista, trouxeram importante contribuição para o nosso entendimento da língua como um fenômeno estreitamente vinculado à vida social dos falantes.

No entanto, desprovidos de uma *teoria social consistente*, esses trabalhos — que tiveram impacto importantíssimo na linguística contemporânea e são merecedores de toda a nossa admiração — não oferecem uma boa visão da dinâmica sociocultural da linguagem. Para se ter essa visão, é necessário recorrer a uma *sociologia da linguagem* que integre as noções de *enunciação* e *ideologia* postuladas por Volóshinov.

Embora, na literatura especializada, os dois termos — sociolinguística e sociologia da linguagem — apareçam frequentemente como sinônimos, também são muitos os autores que insistem em fazer uma distinção entre os campos de atuação de cada uma dessas disciplinas. A principal reprimenda que se dirige à sociolinguística variacionista é a de que, ainda muito apegada ao estruturalismo clássico, ela só recorre ao elemento *socio-* como pretexto para fazer, ao fim e ao cabo, *-linguística*. De fato, essa corrente sociolinguística se dedica primordialmente ao conhecimento *da língua*, à descrição *da língua*, recorrendo a fatores sociais somente na medida em que eles auxiliam nessa descrição. A própria divisão, tão característica dos estudos sociolinguísticos, entre *fatores internos (linguísticos, sistêmicos)* e *fatores externos (sociais)* como vetores da variação e da mudança, entre *variáveis dependentes* e *variáveis independentes*, revela uma dissociação entre entidades que deveriam ser vistas como uma coisa só — ou seja, na sociolinguística variacionista predomina uma concepção *dualista*, que separa o *social* do *linguístico*. No entanto, o social é constitutivo do linguístico, assim como o linguístico é constitutivo do social — não existem "influências" da

sociedade sobre a língua, nem tampouco a língua "reflete" a sociedade: as relações entre esses dois termos, *língua* e *sociedade*, são muitíssimo mais amplas, mais íntimas e mais complexas do que uma mera influência ou um mero reflexo.

Por seu turno, uma verdadeira sociologia da linguagem, como o próprio nome indica, tem de ser antes de tudo um estudo da *sociedade* com foco nas relações que seus membros mantêm entre si por meio da *linguagem*. Aqui o que está em jogo é a língua como *instituição*, como instrumento de *controle social* de uma parcela da sociedade sobre as demais, como *arma* empregada para o exercício da *violência simbólica* de uns cidadãos contra os outros, como palco de *disputas de poder* e de *conflitos entre grupos sociais*, como elemento constitutivo da *identidade de indivíduos e de coletividades*, como *moeda e mercadoria* na economia das trocas linguísticas conforme delineada por Bourdieu... etc.

Não resta dúvida de que os resultados das pesquisas sociolinguísticas são muito importantes para essa sociologia da linguagem e, em muitos aspectos, há mesmo uma sobreposição dos campos de interesse e atuação de cada disciplina. A diferença, porém, é que a sociolinguística variacionista encerra seu trabalho lá onde a sociologia da linguagem (tal como postulamos) mostra que deveria ser o *ponto de partida* (e não de chegada) de uma verdadeira explicação das relações entre uso da língua e dinâmica social. A sociolinguística mostra que existe *variação* e, em certa medida, se detém na explicação das *causas* (linguísticas e/ou sociais) dessa variação, ao passo que a sociologia da linguagem se pergunta: quais as *consequências* (sociais, culturais, éticas, políticas, econômicas, pedagógicas, psicológicas) dessa variação? De que modo a variação linguística é parte *constitutiva* das dinâmicas das relações sociais? Ao contrário da sociolinguística variacionista, que se vale da noção um tanto estática e passiva de que "a língua *reflete* a sociedade", a sociologia da linguagem postula que as línguas *moldam* as sociedades e *são moldadas* por elas, num intrincado jogo de relações, nunca igualitárias.

É por isso que as diferentes modalidades de uso da língua, que nos estudos sociolinguísticos são postas em pé de igualdade, na horizontal, devem aparecer, nessa sociologia da linguagem, numa escala *vertical*, determinada pela hierarquização social dos diferentes falares numa comunidade. Nessa concepção, existem, sim, modos de falar *melhores* e *piores*, *superiores* e *inferiores* — não por alguma característica *gramatical, sistêmica, linguística* intrínseca que eles teriam (e que sabemos que não têm), mas pelos papéis diferentes que são atribuídos a eles no jogo das sempre desiguais relações de poder vigentes na sociedade.

Outro problema que detectamos na sociolinguística variacionista de Labov é sua tentativa de integrar suas investigações no quadro teórico da gramática gerativa. Embora fazendo algumas críticas teóricas e metodológicas à linguística chomskiana, Labov se esforça por formalizar seus achados em regras que recebem notação matemática semelhante às do gerativismo (Labov, 2008: 266):

21 Contração

$$ɔ \rightarrow \langle Ø \rangle \,/\, \left\langle \begin{matrix} +\text{Pro} \\ -\text{cons} \end{matrix} \right\rangle \#\# \left[\frac{\quad}{+\text{Tns}} \right] z \#\# \left\langle \begin{matrix} +\text{Vb} \\ +\text{Fut} \\ -\text{NP} \end{matrix} \right\rangle$$

22 Apagamento

$$z \rightarrow \langle Ø \rangle \,/\, \left\langle \begin{matrix} +\text{Pro} \\ -\text{cons} \end{matrix} \right\rangle \#\# \;\_\_\_\_\; \#\# \left\langle \begin{matrix} +\text{Vb} \\ +\text{Fut} \\ -\text{NP} \end{matrix} \right\rangle$$

Por fim, o apego de Labov ao formalismo estruturalista leva ele a criticar e a rejeitar as premissas das escolas funcionalistas (1994: 547), sempre muito mais voltadas para os fenômenos discursivos, pragmáticos e verdadeiramente sociais do uso da língua.

Deborah Cameron (1995: 18), em suas críticas à sociolinguística variacionista, escreve:

> A tendência nas teorias contemporâneas da linguagem — nas análises pós-estruturalistas/pós-modernistas, bem como na ciência linguística — tem sido a de colocar os sujeitos falantes manipulados (*acted on*) pela linguagem e amplamente incapazes de agir sobre ela por conta própria.

De fato, a partir do momento em que a "língua" se torna um ente com vida e vontade própria, qualquer ação efetiva dos falantes sobre ela parece impossível. No entanto, os membros de toda e qualquer sociedade dispõem da capacidade de agency — palavra usada na sociologia de língua inglesa para definir a capacidade que tem o indivíduo de agir independentemente das estruturas sociais e de fazer suas próprias e livres escolhas. Podemos traduzir *agency* por intervenção, lembrando que *agency* provém do mesmo verbo latino que nos deu agir e ação. Os membros da sociedade não vivem esmagados pelas ideologias dominantes nem pelas estruturas que se filiam a elas: sempre existem brechas e margens de manobra para burlar essas ideologias e estruturas. Como afirma L. E. Alonso (2002: 125), não se pode desprezar

> la *acción* como actividad cotidiana, como capacidad situacional de los actores de operar en un contexto concreto, produciendo sentido en sus actos particulares de habla por medio de procesos de construcción, negociación y resistencia simbólica, incrustados en comunidades culturales de prácticas compartidas, significados cotidianos y actividades rutinarias particulares.

No que diz respeito à intervenção na língua, Cameron (*ibid.*) argumenta:

> Existe um alcance, embora não um alcance ilimitado, para a intervenção efetiva na língua; se negarmos isso em princípio, ou se apagarmos dos registros os seus exemplos específicos, seremos culpados de mistificação.

Ao contrário do que a sociolinguística variacionista deixa implícito (e outras áreas de pesquisa linguística explicitam), as variações e mudanças que ocorrem na língua não ocorrem "naturalmente" — até porque a língua não é um fenômeno exclusivamente "natural" (isto é, biológico). Por seu caráter explicitamente social, cultural e político, a língua é alvo da intervenção consciente ou inconsciente dos falantes. Um fenômeno como a ressurreição do hebraico depois da criação do Estado de Israel em 1948 — ressurreição de uma língua que até então sobrevivia exclusivamente nos textos religiosos tradicionais — é uma clara demonstração disso. Os efeitos do "politicamente correto" na substituição de termos tradicionalmente empregados para se referir a pessoas e eventos, também: os antigos aleijados agora são *deficientes físicos*; as favelas se tornaram *comunidades*; em inglês, os negros já não são *Blacks*, mas *Afro-Americans*, enquanto os brancos se tornaram *Caucasians*... A feminização da linguagem, ocorrida em diversos países, sobretudo de língua inglesa, gerando a chamada *linguagem inclusiva*, é outro exemplo palpável e bem-sucedido de intervenção sobre a língua. E todo o longo processo histórico de *normatização* linguística é decerto a prova mais sólida desse poder de intervenção: mesmo que não empregada por praticamente ninguém, a norma-padrão institucionalizada e prescrita deixa suas marcas nos usos reais de muitos falantes, principalmente com seus efeitos de hipercorreção [▸933].

1.8 Ontogenia e filogenia

O radical de origem grega onto- comporta a ideia de "ser" (de fato, é o particípio passado do verbo *eimi*, 'ser', em grego). Por seu turno, a palavra grega phylē̄ significa "raça, tribo, espécie", e está presente no radical filo- (que não deve ser confundido com outro *filo-*, derivado do verbo grego *philein*, "amar", presente por exemplo em *filosofia*, *filantropia*, *filologia* etc.).

Na linguagem filosófica e científica, a ontogenia se refere ao ser particular, ao indivíduo, enquanto a filogenia se refere ao que é próprio da espécie humana em geral (quanto ao sufixo -genia, consulte [▸284]). Estamos aqui, mais uma vez, diante de uma dicotomia tradicional, a dicotomia *particular* x *universal*, rotulada com outros termos.

As teorias linguísticas que postulam a linguagem como um dado genético, inscrito em nossa herança biológica, se vinculam evidentemente à ideia de filogenia. Não por acaso, são as teorias que hipotetizam a existência de uma gramática *universal* — própria de toda a espécie humana.

Já vimos as críticas de Evans e Levinson e de Christiansen e Chater a esses postulados, bem como as de Michael Tomasello. Esse último autor vai nos guiar a partir de agora numa breve discussão acerca dessa dicotomia.

Baseado em seus muitos experimentos com primatas e também com crianças em fase de aquisição de língua, Tomasello afirma (1999: 94):

> Em discussões sobre cognição humana de um ponto de vista filogenético, a linguagem frequentemente é invocada como uma razão para o caráter único da cognição humana. Mas invocar a linguagem como uma causa evolutiva da cognição humana é como invocar o dinheiro

como uma causa evolutiva da atividade econômica humana. Não há dúvida de que adquirir e usar uma língua natural contribui para (e até transforma) a natureza da cognição humana — assim como o dinheiro transforma a natureza da atividade econômica humana. Mas a linguagem não surgiu do nada. Ela não caiu do espaço sideral na Terra como algum asteroide extraviado nem emergiu como uma bizarra mutação genética desvinculada de outros aspectos da cognição humana e da vida social, apesar das opiniões de alguns teóricos contemporâneos como Chomsky (1980). Assim como o dinheiro é uma instituição social simbolicamente corporificada que emergiu ao longo da história de atividades econômicas existentes anteriormente, a língua natural é uma instituição social simbolicamente corporificada que emergiu ao longo da história de atividades sociocomunicativas existentes anteriormente.

Na linha de Vygotsky e de outros psicólogos culturais, Tomasello afirma que as mais interessantes e importantes façanhas cognitivas da humanidade, como a linguagem e a capacidade de cálculo, exigem tempo histórico e processos para se realizarem. Por isso, sustenta a tese de que essas façanhas cognitivas, essas competências e habilidades requerem um tempo ontogênico significativo, tanto quanto processos, para se realizarem. O autor critica duramente os cientistas cognitivistas por subestimarem a ontogenia em seu papel formativo na criação das formas maduras da cognição humana. Atribui essa postura à supervalorização, por parte dos mesmos cientistas, de um debate que, na sua opinião, é vetusto, caduco, por já ter perdido sua utilidade (se alguma vez foi útil). Esse debate se faz, como temos demonstrado até agora, em torno de polos dicotômicos, de posturas teóricas essencialmente dualistas. A esses dualismos, Tomasello opõe as conquistas teóricas de Charles Darwin (1809-1882):

> As discussões modernas de natureza *vs.* cultura e de inato *vs.* adquirido se estruturam conforme os debates da Europa do século XVIII entre filósofos racionalistas *vs.* empiristas, polemizando acerca da mente humana e das qualidades morais humanas. Esses debates tipificados ocorreram antes que Charles Darwin oferecesse à comunidade científica novos modos de pensar acerca dos processos biológicos. A introdução de modos darwinianos de pensar a filogenia, e o papel da ontogenia na filogenia, deveria ter tornado obsoleto aquele debate. Mas não tornou e, de fato, o surgimento da genética moderna deu a ele vida nova e concretizada na forma de genes *vs.* ambiente. A razão para o debate não ter morrido é o fato dele ser o modo natural de responder a pergunta: o que determina o traço X em seres humanos adultos? Fazer a pergunta dessa maneira, aliás, autoriza as tentativas de quantificar as contribuições relativas dos genes e do ambiente para um dado traço adulto, como "inteligência" [...]. Mas o pensamento darwiniano é um pensamento de processos, no qual não pensamos em categorias de fatores em algum "agora" estático e atemporal. Embora existam processos invariantes como a variação genética e a seleção natural, se perguntarmos como uma dada espécie veio a se tornar o que é agora [...], a resposta é uma narrativa que se desenrola no tempo com diferentes processos agindo em diferentes modos em diferentes pontos ao longo do caminho. (1999: 48-49)

O darwinismo postula (e comprova empiricamente) que a evolução se processa imperceptivelmente pelo acúmulo de transformações ocorridas em indivíduos (portanto, numa escala ontogênica), transformações que, respondendo mais adaptativamente às necessidades de interação entre o indivíduo e seu ambiente, são transmitidas às próximas gerações. Na escala temporal de cada uma dessas gerações, as adaptações prosseguem e se desenvolvem. Desse modo, de ontogênese em ontogênese, as novas

adaptações evolutivas acabam por se constituir em traços característicos daquela espécie — é o que Tomasello chama de "o papel da ontogenia na filogenia".

O chamado *registro fóssil* comprova essa noção: os achados paleontológicos de diferentes espécimes aparentados e separados entre si por lapsos temporais exibem, quando alinhados, o processo de transformação adaptativa de determinados traços biológicos. É assim que descobrimos, por exemplo, que o mais remoto ancestral das baleias era um mamífero terrestre, do tamanho de um gato doméstico, que ficava submerso durante algum tempo, a fim de espreitar suas presas (seu fóssil foi descoberto na Índia, batizado de *Indohyus*, 'porco indiano', e remonta a aproximadamente 48 milhões de anos).

Tomasello que, como já foi dito, se especializou nos experimentos sobre a cognição de chimpanzés e outros primatas, argumenta que a espécie humana tem uma capacidade cognitiva única, herdada biologicamente e que faz ela diferir das outras espécies de primatas: o fato dos seres humanos se "identificarem" mais profundamente com os demais membros de sua espécie do que os outros primatas. Essa identificação não tem nada de misteriosa, mas é simplesmente o processo pelo qual a criança humana entende que as outras pessoas são seres como ela própria. Assim, ela tenta entender as coisas do ponto de vista dos outros. A criança, ao se desenvolver, passa a perceber-se a si mesma como um agente intencional e, mais tarde, como um agente mental. Para o autor, essa diferença cognitiva única

> tem diversos efeitos-cascata porque torna possíveis algumas formas de herança cultural novas e singularmente poderosas. Entender as outras pessoas como agentes intencionais como ela mesma permite tanto (a) processos de sociogênese por meio dos quais múltiplos indivíduos criam colaborativamente artefatos e práticas culturais que acumularam histórias, quanto (b) processos de aprendizagem e internalização cultural por meio dos quais indivíduos em desenvolvimento aprendem a usar e em seguida internalizam aspectos dos produtos colaborativos criados pelos coespecíficos[5]. Isso significa que a maioria (senão todas) das habilidades cognitivas exclusivas da espécie humana não são devidas diretamente a uma herança biológica exclusiva, porém, bem mais, resultam de uma variedade de processos históricos e ontogenéticos que são postos em movimento pela capacidade cognitiva exclusivamente humana, biologicamente herdada (1999: 15).

Compare essas afirmações de Tomasello com a citação feita mais acima de Christiansen e Chater — você verá que esses autores coincidem em grandes linhas.

1.9 A PALAVRA *EVOLUÇÃO*

Parece não haver escapatória: toda teoria complexa, quando "cai na boca do povo", isto é, quando se populariza e se torna artigo fácil no mercado comum do cotidia-

[5] Coespecíficos: indivíduos da mesma espécie.

no, acaba por sofrer drásticas reduções e distorções — quando não são vítimas da síndrome do *"Play it again, Sam"*, frase jamais pronunciada pela personagem Elsie, interpretada por Ingrid Bergman no filme *Casablanca* (de Michael Curtiz, 1942)[6].

Podemos dizer que isso aconteceu com as ideias de Galileu, de Newton, de Marx, de Freud, de Einstein... e também com os trabalhos de Charles Darwin. A principal e mais grave consequência das distorções das ideias darwinianas se encontra no sentido popularmente atribuído à palavra evolução (e também, é claro, ao verbo *evoluir*, ao adjetivo *evoluído* etc.). Que sentido é esse? O sentido de *progresso*, de *aperfeiçoamento*, de *completude* etc. Uma breve pesquisa aleatória confirma facilmente essa distorção:

(1) As escrituras do yoga dizem que uma pessoa evoluída conserva sua raiva por um minuto; uma pessoa comum conserva-a por meia hora e uma pessoa ainda não evoluída conserva sua raiva por um dia e uma noite. Mas uma pessoa cheia de mágoas lembra-se de sua raiva até morrer. (http://www.forumespirita.net/fe/amizade/evoluidos-sentem-raiva-so-por-um-minuto/#ixzz1FCQnvxzB, acesso em 27.2.2011)

(2) Máquina da sorte: como um povo tão evoluído como o japonês pode acreditar em uma tolice dessas?? **Aumenta procura por máquinas da sorte às vésperas do Ano Novo no Japão. Japoneses buscam as máquinas para ver se sua sorte vai mudar com a chegada do novo ano. Procura cresce 80% nesta época do ano.** (http://lineuoateu.blogspot.com/2010/02/maquina-da-sorte-como-um-povo-tao.html, acesso em 27.2.2011)

(3) "DIREITO EVOLUÍDO PRESSUPÕE SOCIEDADE EVOLUÍDA". Sábias palavras do Professor Paulo Nader escritas no livro *Filosofia do direito*, que nos leva a refletir demoradamente acerca das nossas escolhas quanto àqueles que formulam nossas leis. (http://www.artigos.com/artigos/sociais/direito/%93direito-evoluido-pressupoe-sociedade-evoluida%94-6534/artigo/, acesso em 27.2.2011)

Ora, em nenhum escrito de Darwin se encontra a ideia falaciosa de que *evolução* é sinônimo de *aperfeiçoamento, progresso, avanço, desenvolvimento para melhor* etc. Essa distorção, segundo o biólogo Stephen Jay Gould (1941-2002), se prende à tentativa de usar a teoria da evolução e da seleção natural de Darwin para justificar a arrogância da espécie humana de se considerar o ponto máximo dessa evolução, o ser biológico supremo:

> A revolução de Darwin estará completa quando derrubarmos o pedestal da arrogância e ficarmos de posse das implicações simples da evolução quanto à sua tendência não direcional e não previsível — e quando levarmos a sério a tipologia darwiniana, reconhecendo que o *Homo sapiens*, recitando mais uma vez a batida ladainha, é um pequeno ramo, nascido ainda ontem de uma árvore da vida com uma infinidade de ramificações, a qual nunca produziria o mesmo conjunto de galhos se replantada a partir da semente. Nós nos agarramos à palha do progresso (um ramo ideologicamente dessecado) porque ainda não estamos prontos para a revolução de Darwin. Necessitamos do progresso como nossa melhor esperança de conservar a arrogância humana num mundo em evolução. Somente nestes termos podemos entender por que uma tese tão mal formulada e improvável mantém sua poderosa influência sobre nós nos dias de hoje. (Gould, 2001: 49-50)

[6] Também podemos batizar a síndrome de *"Não é um país sério"*, declaração que nunca saiu dos lábios do presidente francês Charles de Gaulle ao se referir ao Brasil. Ou, também, de síndrome do *"Navegar é preciso"*, que tanta gente atribui ao poeta Fernando Pessoa que, na verdade, estava apenas citando um antiquíssimo provérbio latino: *Navigare necesse, vivere non est necesse...*

Segundo o autor, a equiparação de evolução com progresso é mais uma consequência do platonismo que pesa sobre a nossa cultura:

> Na sua célebre analogia da caverna, Platão [...] afirmava que os organismos reais eram apenas sombras na parede da caverna (natureza empírica) — e que um reino ideal de essências deveria existir para lançar as sombras. Poucos de nós defenderiam um platonismo desvairado desse tipo hoje em dia, mas nunca conseguimos abandonar essa opinião característica de que as populações dos indivíduos reais formam um conjunto de acidentes, uma coleção de exemplos fracassados, cada um necessariamente imperfeito e incapaz de se aproximar do ideal além de certo ponto. (p. 66)

O que Gould diz acerca do platonismo na biologia ou no conceito equivocado de evolução que circula tão livremente na sociedade se aplica muito bem (e não há razão para espanto) a muitas concepções de língua e linguagem que circulam nos meios acadêmicos. O fonema, como vimos acima, por exemplo, é um construto platônico, a busca de uma essência, de uma abstrata "unidade sonora" deduzida a partir das realizações imperfeitas produzidas por cada indivíduo. O estruturalismo (e seu ramo mais contemporâneo, o gerativismo) sempre relegou a variação ao submundo da língua, à caverna da linguagem, para se empenhar na busca do mítico sistema, que paira, etéreo, acima do burburinho da vida social. Ora, como se estivesse escrevendo para linguistas, o biólogo Gould afirma:

> Somente admitindo esse platonismo remanescente é que eu posso entender a inversão fatal que tão frequentemente aplicamos ao cálculo das médias. No mundo de Darwin, pós-platônico, a variação sobressai como a realidade fundamental e as médias calculadas se tornam abstrações. Contudo, continuamos a esposar o ponto de vista mais antigo e oposto, e a olhar a variação como um conjunto incoerente de acidentes, valioso principalmente porque nós podemos usar a amplitude para calcular uma média, tomando-a então como a melhor aproximação da essência. Apenas considerando-os como legado de Platão, posso entender os erros comuns sobre tendência que tornam este livro necessário: nossa leitura equivocada em que tomamos a variação contraível e expansível dentro de um sistema como sendo um valor médio (ou extremo) descolando-se numa direção. (p. 66-67; negrito meu)

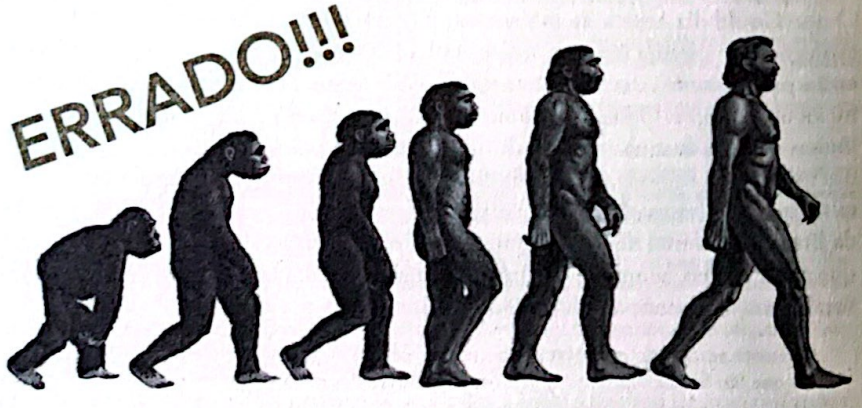
Prossegue o autor afirmando que a revolução darwiniana se completará quando a *essência* for substituída pela *variação* como a categoria central da realidade natural. De fato, no mundo de Platão — e nas teorias estruturalistas e gerativistas que o habitam — a variação era acidental, enquanto as essências apontavam para uma realidade superior. Não é por acaso que, antes do advento da sociolinguística variacionista, se falava de "variação livre" (isto é, acidental) para designar as realizações empíricas das formas ideais da língua, abrigadas no sistema. Na inversão darwiniana,

> prezamos a variação como uma realidade definível (e materialmente concreta), enquanto as médias (nossa aproximação operacional mais chegada às "essências") passam a ser abstrações mentais (*ibid.*).

Stephen J. Gould oferece como contraexemplo da falaciosa sinonímia *evolução* = *progresso* o fato de que as bactérias — os seres vivos mais simples que existem —

"eram no início, são agora e sempre serão os organismos de maior sucesso sobre a Terra" (p. 62). Se a evolução das demais espécies representasse sempre um "progresso", o planeta hoje só deveria ser habitado por organismos extremamente desenvolvidos e "avançados" — no entanto, ainda vivemos na Era da Bactéria, porque esses organismos é que de fato dominam o planeta.

A negação da variação como o motor fundamental da evolução se vê particularmente bem na reproduzidíssima figura da evolução humana, um desenho em tudo e por tudo falso e enganoso.



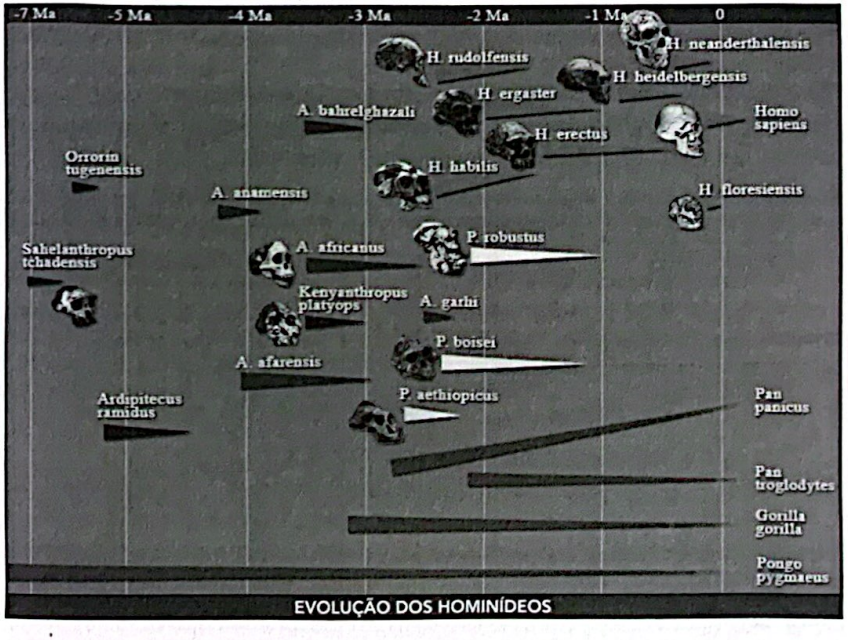
Essa imagem, obsessivamente repetida, banalizada, a ponto de se impregnar na memória coletiva como uma verdade incontestável e absoluta, de fato está repleta de subentendidos (ideológicos) e só serve para nos induzir ao erro. O racismo e o machismo também são evidentes, uma vez que o estágio último, "evoluído", é o homem branco, não uma mulher branca, nem um homem ou uma mulher negra, ameríndia, asiática, polinésia etc. O *Homo sapiens* é, literalmente, um *homem*, branco, inteligente ("sapiente"), enquanto sua fêmea e os coespecíficos de outras "raças" não merecem figurar no esquema da evolução.

Essa figura é tanto mais perigosa porque, além de veicular uma ideia errônea dos princípios da evolução humana, é fácil de compreender e de assimilar. O esquema nos mostra um caminho único da evolução. Cada ancestral tem um descendente, seguido de outro e assim por diante até o homem moderno. Isso faz supor que o homem atual era um "objetivo" da evolução, uma finalidade — talvez um estratagema mental para juntar a teoria darwinista à crença religiosa de que o homem é o dono do mundo, feito à imagem e semelhança de seu Criador... (Criador sempre referido no masculino: o mito bíblico da criação deixa evidente a misoginia impregnada na cultura judaica e legada à cultura cristã, uma vez que Deus cria o homem — macho — à sua imagem e semelhança e, somente depois, retira a mulher de uma costela desse macho, tornando-a desde sempre derivada e dependente dele.)

A evolução, contudo, para desgosto de tantos, não é unidirecional nem é um caminho rumo à "perfeição" (afinal, repetimos, se já existe o ser humano, "perfeito", para quê e por quê ainda existiriam as bactérias?).

O que o registro fóssil realmente nos ensina é que, no lugar da linha reta, o que existiu foi uma intensa e extensa variação na descendência de um provável ancestral comum mais antigo (comum a todos os primatas), que deve ter vivido há cerca de sete milhões de anos. O que se vê no desenho falacioso é um ser muito parecido com um chimpanzé como nosso ancestral mais remoto (o que é no mínimo ridículo, pois deixa entender que o chimpanzé existe desde sempre, em sua forma atual, como se essa espécie não tivesse passado ela também por etapas evolutivas). No entanto, a linhagem que deu origem ao chimpanzé é diferente da que deu origem aos hominídeos. Aqui vemos em ação mais uma deturpação dos postulados de Darwin. Ele nunca disse que "o homem vem do macaco", mas sim que todos os primatas — homens e macacos — descendiam, por linhagens diferentes, de um ancestral comum mais antigo.

Desse ancestral comum mais remoto partiram várias linhagens, como se vê no gráfico abaixo:



Como vamos examinar no capítulo sobre mudança linguística [▸115], a mesma ideia de uma linha reta, partindo do latim e chegando ao português, conduziu a investigação sobre as transformações das línguas, investigação feita durante muito tempo sob o conceito de "língua" como uma entidade dotada de vida própria, independente

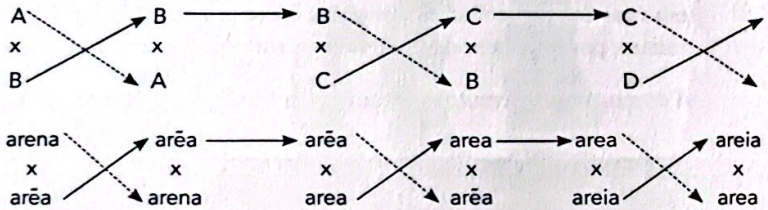
dos mecanismos sociais, culturais, ecológicos, climáticos, políticos etc. que condicionam, tumultuam, organizam e revolucionam as comunidades humanas.

A fórmula tradicional da mudança linguística, como vem estampada nas gramáticas históricas, por exemplo, é semelhante à da imagem da "evolução" do homem: uma linha reta que caminha para a suposta perfeição:

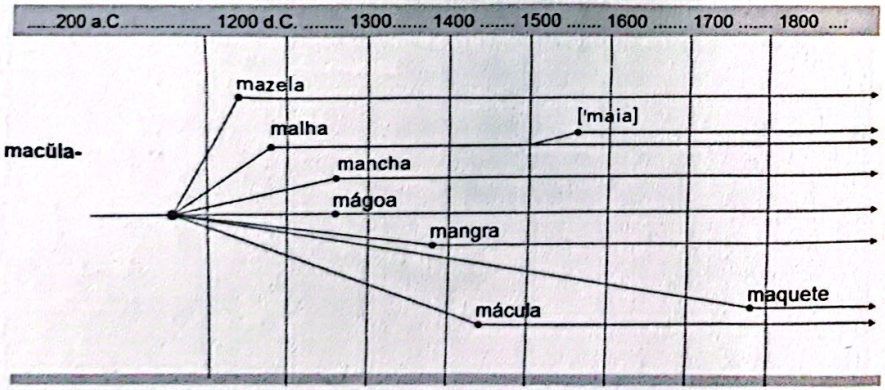
A → B → C → D

arena → arẽa → area → areia

Com a introdução dos fatores sociais no estudo da mudança linguística — principal contribuição da sociolinguística variacionista —, o que se descobriu foi justamente o papel determinante da variação nos processos de mudança:



Além disso, tal como de um mesmo ancestral remoto provieram diversas espécies animais aparentadas, também de um mesmo étimo surgiram diversas linhagens de palavras, cada uma delas com sua própria história evolutiva:



As muitas formas diferentes derivadas de uma ancestral são chamadas *formas divergentes* [▸154].

É preciso alertar que não estamos aqui ressuscitando as velhas hipóteses evolucionistas do século XIX que, numa leitura enviesada dos trabalhos de Darwin, consideravam a língua como um "organismo vivo", que nasce, cresce, vive e morre. O que procuramos mostrar, nessas comparações, é que tanto na biologia quanto na linguística, é imprescindível trabalhar com a variação e abandonar a busca de uma forma "essencial", abstraída das formas empiricamente verificáveis.

A noção de que as línguas mudam rumo à perfeição também esteve presente no pensamento de alguns teóricos do século XIX e permanece, ainda que de forma subjacente, entre os gerativistas, para os quais determinadas mudanças linguísticas poderiam ser motivadas por algo como uma "correção de rumo".

Essa visão da mudança pode ser qualificada de teleológica, isto é, procedente de uma doutrina que identifica a presença de metas, fins ou objetivos últimos que guiam a natureza e a humanidade, considerando a *finalidade* como o princípio explicativo fundamental na organização e nas transformações de todos os seres da realidade. O termo *teleologia* provém do grego *teleios*, 'acabado, completo, concluído'.

Conforme nos relata Carlos Alberto Faraco, essa ideia de completude e aperfeiçoamento (compatível com a falsa sinonímia *evolução* = *progresso*) foi aventada inicialmente pelos importantes pesquisadores do Círculo Linguístico de Praga (1920-1930):

> A mudança, nessa perspectiva, não apenas remodelaria continuamente o sistema, mas faria isso com a finalidade de "corrigir" situações de desequilíbrio. Assim, por exemplo, se uma oposição fonológica se perdeu no processo de mudança, criando, em consequência, palavras homófonas (isto é, palavras com significados diferentes, mas com a mesma estrutura sonora), deveria haver uma nova mudança com a finalidade de "corrigir" essa "incômoda" homofonia. [...] Na base dessas concepções da mudança, está a ideia de que o sistema linguístico tem autonomia e, mais ainda, tem uma dinâmica própria, uma espécie de força intrínseca a presidir sua mudança. (Faraco, 2005: 83, 84)

Ora, acreditar que a língua tem, dentro de si mesma, "em germe", digamos assim, as forças que "causam" sua mudança é resgatar a obsoleta concepção das línguas como organismos naturais, com vida própria.

Se esse pressuposto fosse válido, teríamos de encontrar, em algum tempo ou lugar, línguas "defeituosas", "imperfeitas", incapazes de dar conta das necessidades de interação socioverbal de seus falantes. No entanto,

> nossos registros só conhecem línguas bem-sucedidas: não há registro de língua desaparecida por defeitos internos; não se conhecem línguas aberrantes, nem abortos linguísticos. Nesse sentido, aceitando que as línguas têm funções próprias [...] e considerando que nunca se observou uma língua que, por razões imanentes, não tenha cumprido essas funções, não faz muito sentido considerar qualquer estado de língua ou aspectos de um tal estado como "patológico", "mal adaptado", "desequilibrado". Desde que todas as línguas são, por definição, normais, segue que a anormalidade é uma noção incoerente, ficando difícil sustentar empiricamente teses teleológicas [...] (Faraco, 2005: 90)

Vamos ver a improcedência da noção de "língua primitiva" no próximo capítulo. Tão descabida quanto ela é também a noção de "língua evoluída", no sentido de "língua bem-sucedida", "perfeita, pronta e acabada".

Mais uma vez, o que está aqui em jogo é a arcaica metafísica platônica. A evolução das línguas seria o processo pelo qual elas saem da caverna e avançam rumo à perfeição das formas ideais. Assim, além de teleológico, esse platonismo linguístico é também *teológico*...

Por isso, acompanhando Darwin, podemos dizer, sem susto: qualquer animal "primitivo" estava perfeitamente apto para viver em sua época, em seu hábitat, em seu clima etc. O elefante moderno não é "melhor" nem mais "evoluído" (no sentido de "aperfeiçoado") do que o mastodonte extinto há milhares de anos. Uma forma evoluída de um animal é simplesmente uma forma *diferente*, uma forma que se adaptou às mudanças ocorridas em seu hábitat, no clima global etc.

Da mesma forma, não existe "evolução" (= 'aperfeiçoamento') nas línguas humanas. Os falantes de latim do século IV a.C. pensavam, agiam, interagiam e se comunicavam com a mesma eficácia com que nós, hoje, no Brasil, no século XXI, pensamos, agimos, interagimos e nos comunicamos em português brasileiro. As formas derivadas *malha*, *mancha*, *mágoa* etc. não têm nada de "melhores" ou mais "adequadas" do que *mácula*.

1.10 Mais uma distorção: Comunicar é o que importa

Existe na nossa cultura escolar, no que diz respeito ao ensino de língua, uma ideia muito entranhada e que precisa ser veementemente exposta e combatida. É a noção de que "o que importa é comunicar", de que "se a mensagem foi transmitida, tudo bem", e coisas assim. É fundamental deixar bem claro aqui que não, não e não — essa é uma visão muito pobre e mesquinha do que é a língua e dos mecanismos sociais que a envolvem. Repetir essa ideia é algo extremamente prejudicial para uma boa educação linguística.

Essa ideia é uma deturpação violenta de teorias linguísticas sofisticadas que, lidas pela metade ou só na superfície (quando são lidas), se transformam em conceitos tomados como "verdades científicas" pelos que não se empenham em estudar mais a fundo. E, para piorar, serve de acusação contra os linguistas por parte de pessoas que pretendem, com isso, desqualificar o trabalho dos pesquisadores e tentar preservar a ferro e fogo uma concepção de "língua culta" obtusa, obscura e irreal.

Essas pessoas alegam que para os linguistas "vale tudo", que "o importante é comunicar", que "não é preciso corrigir os alunos", entre outras acusações injustas que não correspondem a nada que linguistas sérios já escreveram ou disseram em público. Para se opor, então, ao que os linguistas jamais disseram, os defensores de uma concepção de língua (e de sociedade) arcaica e pré-científica apregoam o "ensino da gramática" e a inculcação de uma escorregadia "norma culta" como solução para os problemas educacionais brasileiros.

A língua é muito mais do que um simples instrumento de comunicação. Ela é palco de conflitos sociais, de disputas políticas, de propaganda ideológica, de manipulação de consciências, entre muitas e muitas outras coisas. A manipulação social da língua nos leva a votar nessa ou naquela pessoa, a comprar tal ou qual produto, a admitir que determinado evento ocorreu de determinada maneira e não de outra, a aderir a uma ideia, a acreditar nessa ou naquela religião, e por aí vai, e vai longe...

No mercado financeiro, por exemplo, tudo se faz por meio das palavras. Os títulos negociados na Bolsa de Valores não têm existência concreta, são mera abstração, dependem exclusivamente do que se diz ou do que se deixa de dizer: basta lançar um boato sobre uma empresa, dizendo que ela está para falir, e o valor das ações despenca. O que alguns chamam de "invasão" (de terras, por exemplo) outros chamam de "ocupação" (de áreas improdutivas). Onde alguns falam de "terrorismo" outros preferem falar de "revolução". Para os fiéis de determinada religião, certos atos são "pecados", enquanto para os de outra são perfeitamente justificados e bem-vindos. O que o governo americano chamou de "Guerra do Iraque" muitos analistas classificam simplesmente de "invasão", já que os iraquianos não fizeram nada contra os Estados Unidos.

A língua é nossa faculdade mais poderosa, é nosso principal modo de apreensão da realidade e de intervenção nessa mesma realidade. Vivemos mergulhados na linguagem, não conseguimos nos imaginar fora dela — estamos mais imersos na língua do que os peixes na água.

Além disso, a língua é um fator importantíssimo na construção da identidade de cada indivíduo e de cada coletividade. Ela tem um valor simbólico inegável, é moeda de troca, é arame farpado capaz de incluir alguns e excluir muitos outros. É pretexto para exploração, espoliação, discriminação e até mesmo massacres e genocídios, como já vem expresso num conhecido episódio bíblico.

Numa guerra entre duas das tribos de Israel, os galaaditas e os efraimitas, os primeiros se apoderaram dos vaus do Jordão, trechos rasos que podiam ser atravessados a pé. Quando alguém atravessava o rio, os galaaditas mandavam que pronunciasse a palavra *shibboleth* ("espiga"): na variedade linguística dos efraimitas, a palavra era pronunciada *sibboleth*, sem o "chiado" inicial. Quando ouviam essa pronúncia, os galaaditas "*então o matavam nos vaus do Jordão. Caíram naquele tempo quarenta e dois mil homens de Efraim*" (Juízes 12,4-6). Por isso o termo *shibboleth* é usado para

designar qualquer elemento social empregado para discriminar ou mesmo exterminar uma pessoa ou um grupo de pessoas.

Portanto, não se pode admitir essa falácia de que "o importante é comunicar". Abrir a boca para falar é se expor, inevitavelmente, aos julgamentos sociais, positivos e negativos, que configuram a nossa cultura. Falar é comunicar, sim, mas não "transmitir uma mensagem" como ingenuamente se pensa: é comunicar quem somos, de onde viemos, a que comunidade pertencemos, o quanto estamos (ou não) inseridos nos modos de ver, pensar e agir do nosso interlocutor.

Assim, numa sociedade como a brasileira, tradicionalmente excludente e discriminadora, é fundamental que a escola possibilite a seus aprendizes o acesso ao espectro mais amplo possível de modos de expressão, a começar pelo domínio da escrita e da leitura, direito inalienável de qualquer pessoa que viva num país republicano e democrático. A leitura e a escrita, o letramento enfim, abrem as portas de incontáveis *mundos discursivos*, aos quais os aprendizes só vão ter acesso por meio da escolarização institucionalizada.

Por conseguinte, não basta ter o que dizer. É preciso saber dizer o que se tem a dizer: saber usar os múltiplos recursos que a língua oferece para a interação social. E isso é função imprescindível da escola: ensinar a dizer.

1.11 A CONCEPÇÃO DE LÍNGUA DESSA GRAMÁTICA

Todo trabalho acerca da língua tem de definir, inicialmente, a concepção de língua com a qual opera. Proponho uma concepção que sintetiza as mais importantes contribuições das várias (trans)disciplinas que elencamos acima. Essa concepção tem algumas consequências importantes para a elaboração de uma gramática pedagógica como a presente e, na sua esteira, para uma *pedagogia de língua materna*:

1. A gramática de uma língua é sempre emergente, nunca está pronta e acabada, porque seus elementos (fonéticos, morfológicos, sintáticos, semânticos, lexicais) sofrem constantes e ininterruptos processamentos cognitivos da parte dos falantes. Essas operações cognitivo-sociais (abdução, reanálise, metaforização, metonimização, generalizações, restrições etc.) combinadas aos processos sociocognitivos, também ininterruptos, de variação, mudança e contato linguístico impedem a descrição definitiva da gramática de uma língua. Assim sendo, toda descrição é sempre provisória. As categorias, as unidades descritivas, as classes gramaticais etc. não são estáveis e passam o tempo todo por processos de gramaticalização [►170]. Por isso, não tem cabimento "ensinar gramática" na escola, principalmente se esse ensino for sinônimo de memorização acrítica da nomenclatura tradicional e aplicação dessa nomenclatura a frases artificiais, desprovidas de sentido.

2. Sujeita à variação e à mudança ininterruptas, a língua não é um todo homogêneo, um sistema único, mas sim um polissistema, composto das múltiplas variedades (e de suas múltiplas gramáticas) existentes num domínio espacial habitado por uma população que se reconhece falante de uma mesma língua, apesar dessa variação. O estreito vínculo sociocultural, sociogeográfico e sócio-histórico dos falantes com sua língua obriga a considerar que todo idioma nacional é uma língua plena, de modo que não há por que classificar o PB e o português europeu como "variedades" de uma "mesma" língua, uma vez que os falantes brasileiros, ao usar sua língua, estão agindo para a (re)criação e o fortalecimento de sua própria identidade sociocultural e sociolinguística, na mais plena e justificada ignorância do que é e como se constitui a identidade dos portugueses. É no mínimo absurdo pretender que a variedade de um caboclo do interior da Amazônia pertença ao mesmo "sistema" da de um lavrador português do Alentejo, se tudo o mais entre eles é diferente, da cor da pele ao que comem, da latitude em que vivem ao ciclo das estações, do meio ambiente às crenças religiosas, do vestuário aos utensílios etc., e se a intercompreensão entre eles é simplesmente impossível.

3. Sujeita à variação social, essa descrição provisória não tem como escapar ao fato de ser, sempre, a descrição de alguma variedade sociolinguística determinada. Por força das hierarquias de poder e controle social, a variedade que mais tem recebido descrições (e consequentes prescrições) é aquela que historicamente se constituiu e se consolidou entre as classes dominantes da sociedade: urbanas, letradas, quase sempre masculinas, de maior poder econômico e de maior influência política. Essa variedade (que num país das dimensões do Brasil está longe de ser homogênea) se torna objeto de desejo dos que não a têm como variedade própria e consideram seu domínio como mecanismo de ascensão social; e ela também é, de fato, a fonte da norma real, ainda que simbolicamente exista uma outra norma, ideal e prescritiva, definida pelas instâncias oficiais e oficiosas de política linguística e de ensino. Por isso, se é para ensinar alguma norma, que seja, pelo menos, a norma real, o conjunto de variedades realmente empregadas pelos falantes urbanos mais letrados.

4. Nem sincrônica nem diacrônica isoladamente, a língua é um fenômeno pancrônico: num mesmo tempo-espaço social e cultural convivem formas antigas e formas inovadoras, distribuídas desigualmente pelas diversas comunidades de fala que habitam o país de acordo com a história sociolinguística de cada uma. É a dinâmica social que determina a preservação e/ou a extinção de uma forma

linguística e não supostos mecanismos internos à língua agindo por conta própria. A clássica noção de "deriva", isto é, de um movimento autônomo da língua como "coisa-em-si", fica desde logo descartada.

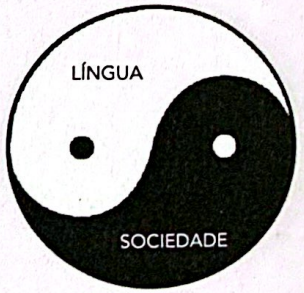
5. A mudança linguística não decorre dessa fictícia deriva "natural" da língua, movida por supostos mecanismos "internos". A mudança decorre sociocognitivamente da interação entre processamentos cognitivos da língua e fenômenos sociais como o contato linguístico, a variação social, a conquista de um povo por outro, a escravização, o poder maior ou menor das forças centrípetas e centrífugas etc. (ver capítulo 4).

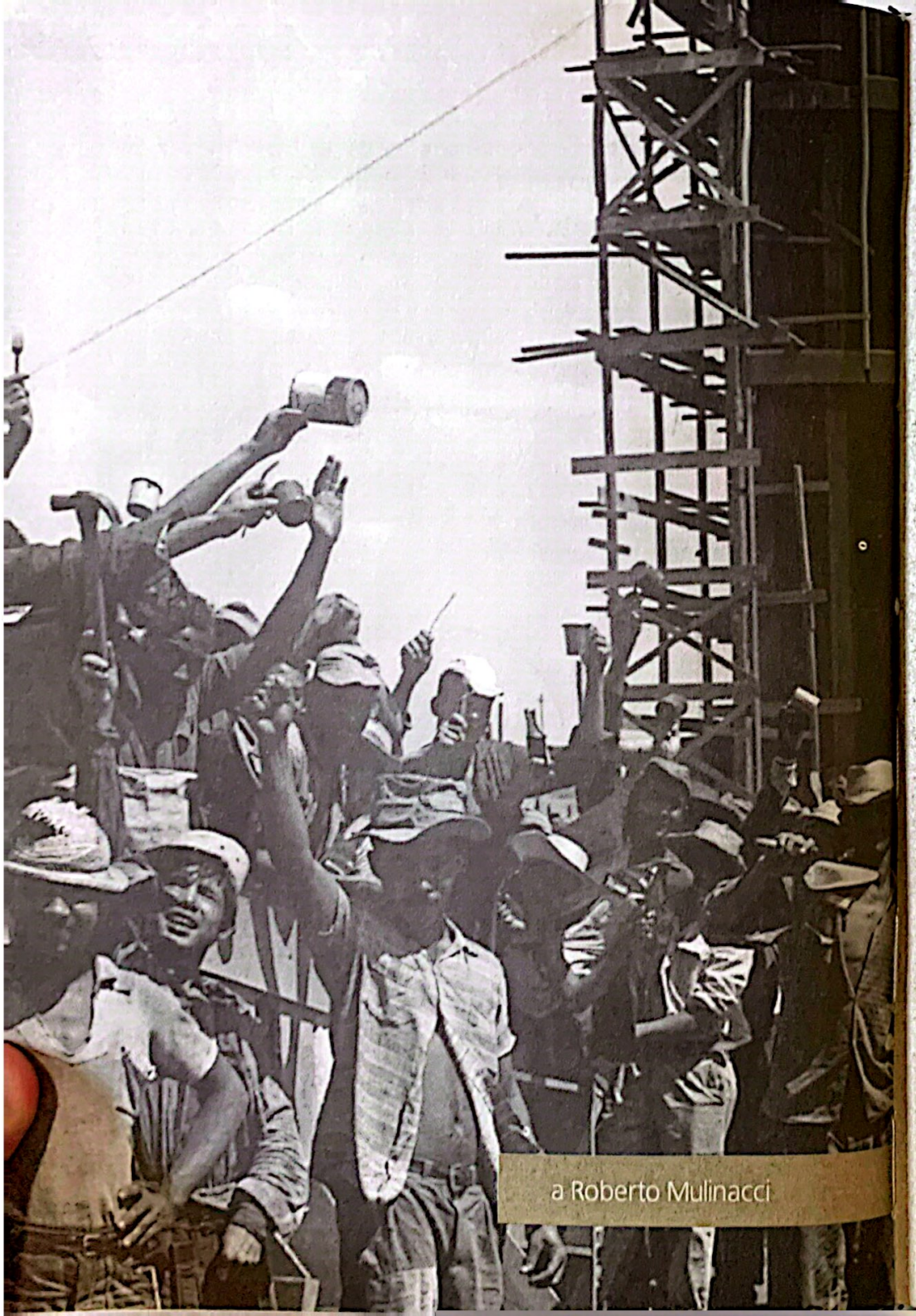
6. Toda manifestação real da língua se dá na forma de textos, falados e/ou escritos, consubstanciados em gêneros discursivos mais ou menos estáveis, culturalmente compartilhados. Por isso, a educação linguística só pode ser eficaz na medida em que partir de textos autênticos, falados e escritos, e nunca de palavras soltas ou de frases artificialmente construídas e descontextualizadas. Nessa perspectiva, desaparece o dualismo certo/errado em prol de uma visão dinâmica dos usos linguísticos e das convenções sociais que os governam e com as quais os falantes podem ou não concordar, sendo perfeitamente legítimo que se rebelem contra elas na medida em que isso fizer parte de seu projeto de vida individual e coletivo. A educação linguística, no entanto, não pode se furtar à tarefa de ensinar o que o falante ainda não sabe, deixando a ele a total liberdade de escolha do que fazer com esse conhecimento adquirido.

7. Língua falada e língua escrita não são dois universos distintos, mas modalidades de uso que se interconectam, se influem mutuamente, se mesclam. A única e verdadeira diferença entre fala e escrita são as condições de produção e de edição de cada modalidade: a fala é produzida e editada em tempo real, no momento mesmo da interação, enquanto a escrita permite correções e emendas antes de ser publicizada. Com as novas tecnologias de comunicação, fala e escrita passam a ter os mesmos recursos semióticos, graças ao desenvolvimento das multimídias. Com isso, os gêneros discursivos se tornam cada vez mais híbridos semioticamente, com interpenetrações profundas entre fala e escrita. A ideia de que é possível livrar a escrita das "interferências" da fala é uma contradição em termos e, como tal, inalcançável, fadada ao fracasso.

8. Os chamados universais linguísticos não são uma herança geneticamente transmitida em nosso DNA. Eles resultam da interação entre as capacidades cognitivas dos seres humanos (que são as mesmas para toda a espécie), as restrições psicofísicas dessas capacidades (o sistema articulatório e respiratório, os limites do processamento mental e da memória etc.) e o mundo físico, com o qual os humanos têm de interagir para dar sentido à existência e manter a coesão social. Uma vez que todos os seres humanos têm de enfrentar os mesmos desafios de conhecimento e experiência com os recursos biológicos de que dispõem, as línguas têm de apresentar um grau de similitude. E uma vez, também, que os ambientes ecológicos e socioculturais em que cada comunidade se desenvolve apresentam diferenças, as línguas igualmente vão apresentar diferenças entre si, mas sempre dentro de um espectro limitado de divergências. A possibilidade de traduzir de uma língua para outra, por mais diferentes que sejam, possibilidade documentada desde os tempos mais remotos da civilização, é a prova empírica cabal da universalidade sociocognitiva dos mecanismos da linguagem humana.

9. Não existe dicotomia entre língua e fala, competência e desempenho, língua-I e língua-E, princípios e parâmetros, forma e substância, significado e significante, pensamento e linguagem, conhecimento e experiência, objetividade e subjetividade etc. O sistema e o uso, a gramática e o discurso, o linguístico e o social compõem um todo indissociável: uma vez separados, a língua simplesmente deixa de ser língua, assim como não se pode dissociar o oxigênio do hidrogênio na água — descombinados, deixam de ser água.





a Roberto Mulinacci

2

o devaneio da língua primitiva

colonialismo, racismo e preconceito linguístico

Absurdo. Aparência. Devaneio. Fábula. Fantasia. Fantasma. Fantasmagoria. Ficção. Ilusão. Lenda. Mito. Superstições. Estereótipos. Quimera...

Quais dessas belas palavras poderíamos usar para tratar de um fenômeno interessantíssimo, comum a praticamente

todas as culturas humanas? O fenômeno de abordar a língua e a linguagem sob pontos de vistas estritamente imaginários, sem correspondência palpável na realidade das coisas?

A linguista francesa Marina Yaguello publicou, em 1988, um livro intitulado *Catalogue des idées reçues sur la langue*. O termo francês *idée reçue* (literalmente: 'ideia recebida') pode ser traduzido por 'preconcepção', 'preconceito', 'prejulgamento' ou qualquer outra palavra que expresse a noção de opinião aceita tacitamente, veiculada pela tradição, pelo senso comum de uma sociedade, sem quase nunca sofrer algum tipo de crítica. Escrevendo em francês para um público de língua francesa, a autora se dedica principalmente às *idées reçues* sobre essa língua que circulam entre os seus falantes. Feitas as devidas adaptações, várias dessas opiniões aceitas sem crítica se encaixariam perfeitamente na descrição do imaginário linguístico dos brasileiros.

Exatos dez anos depois, um grupo de linguistas anglófonos, sob a coordenação de Laurie Bauer e Peter Trudgill, produziu o volume *Language Myths*. São ao todo 28 mitos que, mais uma vez, embora selecionados do imaginário linguístico dos falantes de inglês, também poderiam nos ajudar a entender os fantasmas linguísticos que assombram a cultura brasileira.

Em 1999, sem ainda ter conhecimento da existência dessas duas obras, publiquei um pequeno livro intitulado *Preconceito linguístico: o que é, como se faz*. Nele também recorro à noção de "mito" para tratar de algumas das preconcepções sobre língua e linguagem que circulam em nossa sociedade. Foi com boa surpresa que descobri, pouco depois, que minha iniciativa não era inédita e que os falantes de outras línguas também dispunham de uma discussão crítica acerca das ideias que circulam em suas sociedades quando o assunto é a linguagem humana em geral e línguas particulares.

Idée reçue. Myth. A importância dessas opiniões corriqueiras (e de suas consequências sociais, culturais e políticas) tem sido demonstrada pelo surgimento de campos de investigação delimitados exatamente para estudá-las e entendê-las. Em língua francesa já é comum se falar de *imaginaire linguistique* ('imaginário linguístico'), enquanto em inglês existem diversos trabalhos sobre uma *language ideology* ('ideologia linguística'). Ao contrário da atitude dos primeiros linguistas profissionais — que simplesmente desprezavam e descartavam essas ideias por considerá-las "não científicas" e, portanto, indignas de figurar em seus trabalhos —, a persistência dessas ideias, sua cristalização na mentalidade dos leigos e sua dificílima eliminação chamaram a atenção dos linguistas e de outros cientistas sociais interessados em esclarecer, não só o funcionamento da "língua em si" (uma quimera, ela também, que examinamos no capítulo anterior), mas também a relação dos falantes das línguas com as línguas que eles falam e, mais ainda, com as que eles não falam. A sociolinguística, a análise do discurso, a pragmática, a etnografia da comunicação, a sociologia da linguagem, a antropologia linguística... todas essas disciplinas reconhecem a pertinência do estudo dessas preconcepções, talvez a única forma de melhor denunciá-las e ajudar em sua eliminação.

A linguista inglesa Deborah Cameron, criadora da noção de *higiene verbal* (1995), acusa diretamente os linguistas que tratam da "língua em si" de serem os principais responsáveis pela manutenção das ideias preconcebidas sobre a linguagem. Por higiene verbal a autora entende um traço universal do comportamento dos seres humanos: nós não só falamos línguas como também *falamos das línguas* que falamos ou que os outros falam. Torcendo o nariz para essas ideias e retirando-se do debate público para se dedicar ao exame exclusivo do "sistema" (uma entidade que também é mítica, por mais que se pretenda qualificá-la de "científica"), a maior parte dos linguistas não tem desempenhado o papel político que poderiam muito bem desempenhar: mostrar, de forma cabal e convincente, que aquelas ideias não têm fundamento na realidade da língua e que, aparentemente inofensivas, são de fato extremamente prejudiciais para o bom convívio das diferenças entre as línguas e entre as muitas variedades de uma mesma língua. Esses mitos também são muito prejudiciais para a educação linguística.

Infelizmente, porém, boa parte da responsabilidade pela divulgação e manutenção desses mitos e preconceitos tem cabido à escola e às ideologias de ensino de língua que configuram ainda (tristemente) em grande medida o trabalho pedagógico — e não somente no Brasil, diga-se de saída. De fato, o linguista e educador britânico Michael Stubbs escreveu:

> Toda a área da língua na educação está impregnada de superstições, mitos e estereótipos, muitos dos quais têm persistido por séculos e, às vezes, com distorções deliberadas dos fatos linguísticos e pedagógicos por parte da mídia. [...] Existe uma generalizada incapacidade de encarar a língua como um objeto de reflexão crítica: ela quase sempre é considerada óbvia demais para ser digna de estudo, ou misteriosa demais para ser explicada. As opiniões prescritivas sobre "padrões" parecem simplesmente incontestáveis, e os "problemas" da língua parecem merecer um tratamento fácil e superficial. (2002: 157)

Se a escola tem sido a responsável por esse grave problema, chegou então a hora dessa mesma escola, numa sociedade democrática e respeitosa da pluralidade cultural e linguística, examinar os mitos que a sustentam e abandoná-los o quanto antes. É impensável, em pleno século XXI, que ideias tão retrógradas continuem circulando com toda a tranquilidade em nosso ambiente escolar.

Vamos pôr lado a lado as "ideias preconcebidas" do livro de Yaguello e os "mitos" dos autores britânicos[1]:

[1] Os mitos que abordei em meu livro de 1999 foram: (1) o português falado no Brasil apresenta uma unidade impressionante; (2) brasileiro não sabe português (só os portugueses sabem); (3) português é muito difícil; (4) as pessoas sem instrução falam errado/feio; (5) o lugar onde melhor se fala é o Maranhão; (6) o certo é falar assim porque se escreve assim; (7) é preciso saber gramática para falar e escrever bem; (8) o domínio da norma-padrão é um instrumento de ascensão social.

YAGUELLO (1988)	BAUER & TRUDGILL (1998)
1. "alguns povos são mais dotados para falar línguas do que outros",	1. Não se deve permitir que o significado das palavras varie ou mude.
2. "existem sotaques vulgares e sotaques elegantes",	2. Algumas línguas simplesmente não servem.
3. "os negros não conseguem pronunciar o *r*",	3. A mídia está arruinando o inglês.
4. "existem línguas ricas e línguas pobres; línguas simples e línguas complexas",	4. O francês é uma língua lógica.
5. "as classes gramaticais têm que ser as mesmas em todas as línguas",	5. A ortografia do inglês é kattastrófika.
6. "as línguas sem escrita não têm gramática",	6. As mulheres falam demais.
7. "a gente nunca esquece a língua materna",	7. Algumas línguas são mais difíceis que outras.
8. "se não está no dicionário, não existe", "nossa língua está em decadência",	8. As crianças não conseguem mais falar ou escrever direito.
9. "o que não é claro não é francês",	9. Nos Apalaches eles falam como Shakespeare.
10. "o francês é a língua mais difícil do mundo",	10. Algumas línguas não têm gramática.
11. "existem línguas primitivas",	11. Italiano é bonito, alemão é feio.
12. "existem línguas bonitas e línguas feias".	12. Gramática ruim é desleixo.
	13. As crianças negras são deficientes verbais.
	14. Negativa dupla é ilógica.
	15. A TV faz as pessoas falarem igual.
	16. Não se deve dizer *it's me* porque me é acusativo.
	17. Eles falam mesmo um inglês ruim no Sul e em Nova York.
	18. Algumas línguas são faladas mais depressa do que outras
	19. Os aborígines falam uma língua primitiva.
	20. Todo mundo tem sotaque menos eu.
	21. Os americanos estão arruinando o inglês.

Como os próprios títulos dos livros já indicam, todas essas ideias são *falsas* — quimeras, ilusões, enganos, fantasias etc. — porque não correspondem a nada que os pesquisadores têm descoberto sobre as línguas humanas desde que começaram a se debruçar sobre elas com um instrumental de análise mais acurado e, igualmente importante, movidos por uma ideologia de sinal contrário à que governa essas noções antigas — a ideologia de que as opiniões tradicionais sobre a língua estão no mesmo plano dos demais preconceitos que circulam em qualquer sociedade humana: "as mulheres são seres inferiores", "os índios são atrasados mentalmente", "os negros são naturalmente violentos", "os homossexuais são doentes maníacos por sexo", "as religiões politeístas lidam com magia negra", "o islamismo é uma religião do mal", "os judeus são sovinas", "os árabes são terroristas", "os japoneses são trabalhadores", entre tantas e tantas outras.

Por isso, não vamos nos dedicar aqui a analisar todas essas ideias. É preciso apenas reforçar o apelo para que — em sala de aula — a professora não se deixe levar por tais fantasias e, além disso, também participe da militância para denunciá-las e combatê-las.

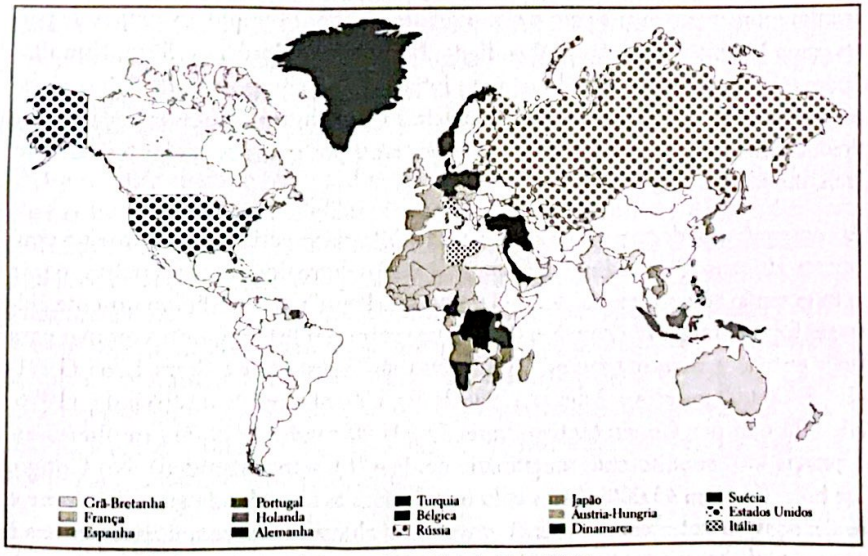
Uma ideia, porém, merece nossa atenção particular — a de que existiriam línguas "primitivas". A análise atenta desse mito nos ajuda a compreender uma série de

outras questões históricas, políticas e ideológicas. Desmontar esse mito é importante porque, há muito tempo, ele vem sendo associado perversamente a outras manifestações igualmente preconceituosas, usadas como justificativa para a opressão, o massacre e a escravização dos povos.

2.1 O COLONIALISMO E SEU FESTIVAL DE HORRORES

O final do século XIX foi marcado por uma política extremamente agressiva das potências coloniais europeias. Depois da independência dos Estados Unidos em 1776 e dos países latino-americanos na primeira metade do século XIX, as potências europeias, para recuperar as riquezas perdidas, se lançaram numa empreitada que passou à história com o nome de *Partilha da África* ou *Corrida pela África*. Até então, a exploração do continente africano, de seu povo e de suas riquezas era feita de modo mais ou menos "informal", levada a cabo sobretudo por comerciantes, traficantes, contrabandistas etc.

A partir de 1880, no entanto, os governos europeus se engajaram, de forma institucionalizada e explícita, na disputa pelo continente. Inglaterra, França, Portugal, Espanha, Itália, Alemanha e Bélgica ocuparam imensas áreas da África, delimitando fronteiras, criando instituições oficiais, enviando exércitos, missionários religiosos, comerciantes, professores, médicos, abrindo hospitais, escolas, quartéis etc.



[A situação do mundo colonial em 1914.]

Em toda a África, somente dois países se mantiveram independentes: a Libéria e a Abissínia (atual Etiópia).

A Libéria tinha sido criada em 1847 por ex-escravos que aceitaram a proposta de grupos e instituições americanos de serem "devolvidos" à África, onde poderiam viver mais "livremente" e mais "integrados" à cultura "original" do que nos Estados Unidos, país em que jamais poderiam (e até hoje praticamente não podem) se integrar à sociedade dominante branca. Daí o nome *Libéria*, derivado do latim *liber*, 'livre'. Na realidade, a criação da Libéria era só o primeiro empreendimento da ACS (American Colonization Society — Sociedade Colonial Americana), que pretendia adquirir, como as demais potências, áreas coloniais na África. No entanto, esse empreendimento não teve sucesso.

A Abissínia, por seu turno, era um império muito antigo e, principalmente, era (e é) o único país africano em que o cristianismo se impôs muito cedo, fazendo surgir a Igreja Copta. A dinastia etíope remontava ao século II a.C. (e só foi deposta em 1974) e a cristianização do país se iniciou em 316 d.C. Além disso, não apresentava nenhum tipo especial de riqueza natural que merecesse a avidez das potências europeias.

A Partilha da África (e também do Oriente Médio) fica evidente quando observamos o traçado das fronteiras dos países que surgiram após os movimentos de independência do continente. Em muitos deles, as fronteiras são linhas retas, como se tivessem sido traçadas com réguas e esquadros. Na verdade, elas *foram mesmo* traçadas com réguas e esquadros — diante da possibilidade de eclodirem guerras entre as potências coloniais pela disputa daquelas áreas, elas se reuniram e fizeram acordos sobre quem ficaria com o quê. Procure ver, por exemplo, os mapas de países como Mauritânia, Argélia, Líbia, Egito, Mali, Iraque, Jordânia, Síria, Kuwait... Todos têm linhas retas, quando não são inteiramente compostos de linhas retas, ao contrário de outras nações, cujas fronteiras foram historicamente traçadas por desdobramentos políticos internos e externos e/ou por grandes acidentes naturais (rios, montanhas, mares etc.).

As consequências dessa partilha autoritária e arbitrária se verificam até hoje no continente africano. Povos e línguas ficaram divididos entre dois ou mais países, o que vem gerando sempre muitas tensões e guerras. Populações tradicionalmente inimigas foram obrigadas a conviver dentro das mesmas fronteiras, outro motivo para mais guerras e mais massacres. Os números são assustadores: Serra Leoa (1991-2000: 50.000 mortos); Libéria (1998-2003: 200.000 mortos); Ruanda (1996: 850.000 mortos); Congo (antigo Zaire: 5.400.000 mortos; 200.000 mulheres estupradas — o conflito mais sanguinário desde a II Guerra Mundial). No Congo, até hoje, ocorrem 45.000 mortes todo mês devidas às sequelas da guerra: a fome e as doenças, sobretudo a AIDS, sendo que a mais alta porcentagem dessas vítimas é composta de crianças.

A conquista da África foi, evidentemente, um banho de sangue. Com a Revolução Industrial, os europeus tinham desenvolvido uma indústria armamentista poderosa, moderna e eficiente, contra a qual as armas tradicionais dos povos africanos não puderam oferecer resistência. Para mencionar apenas dois horrores cometidos pelos europeus nesse período, podemos citar o massacre dos povos herero e namáqua e as exibições de seres humanos como atrações de circos e zoológicos.

Na virada do século XIX para o XX, a Alemanha tinha o controle da região que hoje constitui a Namíbia. O governo alemão tentou explicitamente exterminar dois povos locais, os hereros e os namáquas, que ofereciam forte resistência à conquista alemã. No total cerca de 65.000 hereros (80% da população total) e 10.000 namáquas (50% da população) foram mortos entre 1904 e 1907. Os métodos mais empregados foram a morte pela fome e o envenenamento dos poços de água dessas populações enquanto estavam encurraladas no deserto da Namíbia. O general alemão Lothar von Trotha, responsável pelo genocídio, escreveu sem rodeios sobre o povo herero:

> Estou convencido de que essa nação como tal deve ser aniquilada ou, se isso não for possível por medidas táticas, deve ser expulsa do país... Isso será possível se os poços de água de Grootfontein até Gobabis forem ocupados. O constante movimento de nossas tropas nos permitirá encontrar os pequenos grupos da nação que tiverem recuado e destruí-los gradualmente.

Repetiram-se, portanto, na África, em pleno século XX, as mesmas práticas genocidas que caracterizaram, a partir do século XVI, a conquista das Américas por espanhóis, portugueses e ingleses. (Práticas que, em certa medida, prosseguem até hoje no Brasil, com assassinatos sistemáticos de indígenas por grileiros, posseiros e mineradores.) O massacre dos hereros e dos namáquas é considerado hoje o primeiro genocídio do século XX, que viria a conhecer muitos outros e ainda mais monstruosos. (Por sua vez, o século XXI também já teve seu primeiro genocídio na região de Darfur, no Sudão, onde forças ligadas ao próprio governo sudanês são suspeitas de ter matado mais de 400.000 pessoas.)

Outra monstruosidade praticada nesse período (até os anos 1930) foram as "exposições coloniais" em que indivíduos trazidos da África, do Ártico e da Polinésia eram exibidos como atrações exóticas em "zoológicos humanos" de diversas cidades europeias e também norte-americanas. Samoanos, tuaregues, núbios, pigmeus, hotentotes, malgaxes, inuits ("esquimós") eram mantidos em jardins botânicos e jardins zoológicos, muitas vezes em jaulas, na mesma seção de macacos e outros animais selvagens. Novamente, repetiam-se práticas comuns durante a "descoberta" da América, em que indígenas de todo o continente foram levados à Europa como "presentes" para reis e rainhas ou "objetos exóticos", ao lado de animais nunca vistos antes como papagaios, macacos, jacarés etc.

Essas exposições tinham, além de sua função de propaganda junto às populações dos países europeus, um amplo apoio "científico". No século XIX, o racismo ganhou

ares de "ciência" com o desenvolvimento da eugenia, uma teoria pseudocientífica que tentava mostrar a superioridade da "raça" branca sobre as demais "raças" humanas — e até mesmo dentro da "raça" branca havia uma hierarquia de superioridade: os nórdicos acima dos latinos e outros povos mediterrâneos, por exemplo.

Entre os cientistas e pensadores que defenderam a *eugenia*, se destacam o francês Arthur de Gobineau (1816-1882), que esteve no Brasil e ficou horrorizado com a nossa "mistura de raças", e o zoólogo e geólogo suíço Louis Agassiz (1807-1873), que também veio fazer pesquisas no Brasil e escreveu:

> Aqueles que põem em dúvida os efeitos perniciosos da mistura de raças e são levados, por falsa filantropia, a romper todas as barreiras colocadas entre elas deveriam vir ao Brasil.

Com base nos princípios eugenistas, a colonização se justificava plenamente: era preciso levar a "civilização" aos povos "selvagens", fazê-los abandonar seus costumes "bárbaros", impor a eles a religião cristã, de modo que deixassem de adorar seus "ídolos" e "demônios" e de praticar "feitiços" e "magia negra". Também era preciso impor a eles as "línguas civilizadas" para que deixassem de falar por meio de grunhidos e urros. Essa ideologia se impregnou profundamente na mentalidade europeia da época, fazendo muita gente bem-intencionada acreditar na necessidade de "civilizar" os outros povos e até mesmo levando muitos a se engajar nessa empresa (missionários religiosos, médicos, professores etc.).

2.2 A PSEUDOCIÊNCIA DAS LÍNGUAS PRIMITIVAS

Nesse grande projeto ideológico colonialista, a ciência linguística da época deu sua contribuição, alegando a existência de "línguas primitivas". O caráter primitivo dessas línguas era considerado óbvio e natural: como é que povos atrasados, inferiores, intelectualmente retardados, selvagens etc. poderiam falar, a não ser por meio de grunhidos, rangidos, frases toscas, vocabulário pobre etc.? A própria condição de língua era negada a esses modos de falar, que recebiam, quando muito, o rótulo de "dialetos". (Entre 2000 e 2001 a Rede Globo de Televisão exibiu a telenovela Uga-Uga — um título que já diz tudo sobre a ideologia retrógrada dos autores e produtores da obra.)

Com o avanço das ideias progressistas, com as revoluções sociais e políticas ocorridas desde aquela época até hoje, com a democratização das sociedades europeias e americanas, com a conquista e ampliação dos direitos humanos, o racismo "científico" foi denunciado e demolido pelas versões contemporâneas da linguística, da sociologia, da antropologia, da psicologia e outras ciências. Infelizmente, porém, o efeito da propaganda colonialista persistiu e *persiste, muito arraigado* nas culturas populares e até mesmo entre pessoas *que se acham* muito cultas.

O linguista Sírio Possenti relata, por exemplo, numa de suas crônicas, um episódio ocorrido no programa de entrevistas de Jô Soares na televisão:

> Um dia desses, os entrevistados eram Martinho da Vila e o presidente de uma associação de magistrados. [...] A conversa com Martinho ia bem, até que Jô perguntou sobre seu conhecimento de línguas africanas, já que de alguns dos discos de Martinho participam músicos angolanos cantando músicas nativas. Martinho disse o óbvio: que, tendo estado na África várias vezes, mesmo em temporadas curtas, aprendeu um pouco. Não conhece as línguas, mas se vira (e acrescentou que o mesmo ocorre com relação ao francês, o que mostra que ele é normal). Mas Jô o interrompeu para comentar que se pode aprender as línguas africanas, mesmo em pequenas temporadas, porque elas têm poucas palavras. E botou para funcionar suas leituras de almanaque. Informou que em suaíli as palavras querem dizer muitas coisas. E deu como exemplo certa palavra que pode ser empregada em várias situações. Decidi dormir, perdi a entrevista com o magistrado. Achei que não suportaria uma lição de direito constitucional do mesmo nível.
> Disse que Jô acionou suas leituras de almanaque, mas a coisa é mais grave do que isso: trata-se de grosseiro preconceito linguístico e cultural. Se a gente abre um dicionário [...], a coisa mais interessante que se pode descobrir é que todas as palavras têm muitos sentidos, que todas as línguas são como o suaíli, ou o suaíli é como todas as línguas. (Possenti, 2009: 18-19)

Possenti dá como exemplo o verbo *ligar*. Uma consulta rápida ao dicionário eletrônico *Houaiss* nos revela 16 acepções diferentes de *ligar*, que vão desde "unir, prender, atar (algo ou alguém) com um liame ou laço, para reunir dois ou mais elementos da mesma natureza, ou assegurar a coesão das partes de um todo, ou impedir uma ação ou um movimento" até "pôr em marcha (motor)", num aparente paradoxo: como é que o mesmo verbo que significa "impedir uma ação ou um movimento" também significa "pôr em marcha (motor)"? Seria o português uma língua "primitiva"?

É óbvio que não. Como diz Possenti, a polissemia, a capacidade que as palavras têm de comportar múltiplos sentidos, é uma característica de absolutamente toda e qualquer língua humana.

O que caracterizaria uma língua "primitiva"? A pobreza de vocabulário já vimos que não existe. Seria, então, a pobreza na gramática, nos recursos sintáticos do idioma? Uma comparação rápida entre o latim clássico e o latim vulgar e, mais adiante, entre o latim vulgar e as línguas românicas mostra as inúmeras "perdas" que ocorreram no processo. No entanto, é claro que também ocorreram muitos "ganhos", porque afinal toda língua humana tem que dar conta de expressar uma série de categorias sociocognitivas sem as quais o conhecimento e a comunicação seriam impossíveis (tempo, lugar, espaço, pessoa, objeto, visão, movimento, evento etc.). No entanto, pelos princípios da mudança linguística que vamos examinar no capítulo 4 — economia linguística, gramaticalização, analogia —, essas categorias passam a ser expressas de outras maneiras. Se as línguas românicas abandonaram as declinações do latim, as funções sintáticas passaram a ser exercidas pela ordem dos termos na oração e pelo uso das preposições. O latim não apresentava artigos nem pronomes de não-pessoa, lacunas que foram supridas em todas as línguas românicas.

Se existissem línguas "primitivas" também poderíamos falar de línguas "desenvolvidas". O que caracterizaria essas línguas? Difícil (impossível) responder. Se o paralelo for feito entre desenvolvimento tecnológico e "desenvolvimento" linguístico — como muita gente, equivocadamente, faz — poderíamos dizer que as línguas mais "desenvolvidas" seriam as línguas faladas pelos povos mais desenvolvidos do ponto de vista científico, tecnológico, econômico etc. Ora, qual é a língua dos dias de hoje que corresponde melhor a essa definição senão o inglês?

O inglês, no entanto, é uma língua extremamente simples, do ponto de vista morfossintático. Observe, por exemplo, quantas traduções possíveis existem, em PB, para uma única forma verbal inglesa:

YOU COULD	você pôde /	vocês	puderam
	podia		podiam
	poderia		poderiam
	pudesse		pudessem
	puder		puderem

O inglês não tem, como todas as demais línguas indo-europeias, marcas de gênero para os substantivos e adjetivos — por seu turno, as línguas do grupo germânico, de onde provém o inglês, classificam suas palavras em *masculinas*, *femininas* e *neutras*. Os determinantes em inglês, ao contrário do que acontece em tantas outras línguas, não trazem marcas de gênero nem de número: *the* ('o, a, os, as'), *my* ('meu, minha, meus, minhas'), *all* ('todo, toda, todos, todas, tudo') etc. Já em alemão, os artigos definidos assumem diversas formas, a depender do gênero, do número e do caso sintático do termo determinado: *der, die, das, den, dem, des* etc.

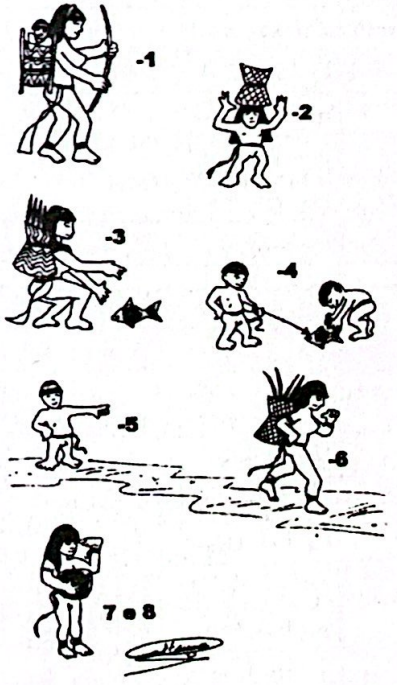
Ao longo do tempo, portanto, o inglês sofreu drásticas simplificações morfossintáticas — a morfologia verbal, simplíssima, é o melhor exemplo disso: enquanto, para aprender bem o PB, um estrangeiro precisa conhecer mais de dez tempos verbais com suas respectivas desinências, mais as marcas de pessoa e número, mais o particípio passado (que tem marcas de pessoa e número) — e isso só para os verbos regulares, que não trazem mudança de vogal no radical nem mudança de radical —, para aprender inglês basta aprender o infinitivo e o particípio passado e saber que a não-pessoa do singular ganha um *-s*: *love, loved, loves*. Será que o inglês é uma língua pobre, primitiva?

Bobagem. Tudo o que os falantes do antigo anglo-saxão (o antepassado remoto do inglês) conseguiam expressar com suas morfologias verbais e nominais complexas, os falantes de inglês contemporâneos também conseguem, e até mais, porque o mundo se transformou radicalmente nesses últimos mil e quinhentos anos, de modo que foi preciso criar um vastíssimo vocabulário e uma ampla rede de relações lexicogramaticais para dar conta de expressar o mundo moderno.

Por outro lado, quando examinamos uma língua falada por uma população tradicional, que se mantém vinculada a seu território de origem e à sua cultura ancestral, ficamos surpresos com a extrema complexidade que muitas delas apresentam. A língua tuiuca, por exemplo, falada no extremo norte do Brasil e na Colômbia, tem toda uma morfologia própria para expressar, no verbo, a origem da informação que está sendo transmitida (uma conjugação epistêmica, por assim dizer) (Martin, 2003: 88):

diiga apé-wi	Ele jogou: eu vi
diiga apé-ti	Ele jogou: eu ouvi
diiga apé-yi	Ele jogou: tenho indícios disso (por exemplo, as marcas de chuteiras no campo)
diiga apé-'yigi	Ele jogou: me disseram
diiga apé-hiyi	Ele jogou: pode se supor (por exemplo, porque ele tem o hábito de jogar na sexta-feira e hoje é sexta-feira)

Nenhuma língua indo-europeia apresenta morfologia semelhante.



1. **ã kunumĩ** – *este menino* – próximo e fora da vista (Veja que o menino está nas costas da mulher.)
2. **ang mba'e** – *esta coisa* – próxima e fora da vista (Veja que o cesto está na cabeça da menina, que não o vê.)
3. **kó pirá** – *este peixe* – próximo e visível
4. **ebokûeĩ pirá** – *esse peixe* – perto da pessoa com quem se fala e visível
5. **kûeĩ kunhã** – *aquela mulher* – distante e visível
6. **akûeĩ kunumĩ** – *aquele menino* – distante, ausente e fora da vista (Veja que a mulher está de costas para o menino e não o vê.)
7. **kó uru** – *esta vasilha* – próxima e visível (A vasilha está na mão da mulher.)
8. **aĩpó nhe'enga** – *aquela voz* – de alguém que se ouve mas não se vê

Por sua vez, o tupi antigo, como vemos na figura da página anterior *[▸91]*, apresenta um sistema de demonstrativos que está longe de ser "simples" ou "primitivo". Segundo nos informa Navarro (2006: 36, fonte da imagem que reproduzimos), os demonstrativos do tupi se distinguem "conforme a *proximidade* e a *visibilidade* (i. e., variam, dependendo do fato de as coisas ou as pessoas serem vistas ou não").

Enquanto em PB e nas demais línguas românicas as palavras são classificadas em masculinas e femininas, as línguas do grupo banto classificam suas palavras num número muito maior de "gêneros". O chamado banto comum, o ancestral mais remoto das línguas do grupo, apresentava nada menos do que 22 classes. O suaíli, língua banta amplamente difundida na África subsaariana como meio de comunicação entre diferentes povos, apresenta 17 classes nominais. Cada uma dessas classes tem prefixos próprios que são agregados aos substantivos, adjetivos, verbos, sujeitos e objetos, num sistema de concordância extremamente complexo:

CAMPO SEMÂNTICO	NOME -CON, -VOG	SUJ	OBJ	ADJETIVO* -CONS, -I, -E
1 pessoa	m-, mw-	a-	m-	m-, mwi-, mwe-
2 pessoas	wa-, w-	wa-	wa-	wa-, we-, we-
3 árvore	m-	u-		m-, mwi-, mwe-
4 árvores	mi	i-		mi-, mi-, mye-
5 grupo, AUM	ji-, ø, j-	li-		ji-/ø, ji-, je-
6 grupos, AUM	ma-	ya-		ma-, mi-, me-
7 ferramenta, DIM	ki-, ch-	ki-		ki-, ki-, che-
8 ferramentas DIM	vi-, vy-	vi-		vi-, vi-, vye-
9 animais, 'outros' 10 empréstimos	N-	i-		N-, nyi-, nye-
		zi-		
11 extensão	u-, w-/uw-	u-		m-, mwi-, mwe-
12 plural de 11	N-	zi-		N-, nyi-, nye-
13 abstração	u-, w-/uw-	u-		m-, mwi-, mwe- *ou* u-, wi-, we-
14 infinitivos	ku-, kw-	ku-		ku-, kwi-, kwe-
15 posição	-ni, mahali-	pa-		pa-, pi-, pe-
16 direção, em torno	-ni	ku-		ku-, kwi-, kwe-
17 dentro, junto	-ni	mu-		mi-, mwi-, mwe

[(*) Em suaíli, os adjetivos começam sempre por consoante ou pelas vogais *i* e *e*].

Exemplos desse sistema de concordância:

(l) yu-le m-tu m-moja m-refu a-li y-e- ki-soma ki-le ki-tabu ki-refu.
'Aquela uma-pessoa alta que lia aquele longo livro'

(l') wa-le wa-tu wa-wili wa-refu wa-li (w)-o vi-soma vi-le vi-tabu vi-refu.
'Aquelas duas pessoas altas que liam aqueles longos livros'

A essa mesma família linguística pertencem as línguas africanas que mais influenciaram o PB: o quimbundo, o quicongo e o umbundo. Todas elas apresentam o mesmo sistema de concordância, com variação do número de classes de uma para a outra. Fica evidente, de saída, que a ideia, tão divulgada durante séculos no Brasil, de que os negros "deturparam" o português porque suas línguas eram "primitivas", "toscas" e "simplórias" não tem a menor sustentação, a não ser no preconceito racial e na ideologia colonialista. O sistema de concordância das línguas bantas é muito mais sofisticado do que o das línguas românicas que, conforme sabemos, já representam uma simplificação do sistema de concordâncias do latim.

Um exemplo final nos vem da Austrália. Existem muitas línguas diferentes naquele país — além do inglês oficial —, todas chamadas genericamente de "línguas aborígines" que, no entanto, pertencem a famílias distintas, sendo mesmo incompreensíveis entre si. Na época da chegada dos colonos ingleses, calcula-se que haveria entre 200 e 300 línguas nativas, hoje reduzidas a apenas 70. (O mesmo aconteceu no Brasil: das mil línguas que existiriam por aqui na época da invasão portuguesa, hoje restam pouco mais de 180.) Numa dessas línguas, o kunwinjku, existem nomes específicos para animais que, em inglês, apresentam um único nome — todos da família dos macrópodes (cangurus etc.):

NOME INGLÊS E NOME CIENTÍFICO	KUNWINJKU		
	MACHO	FÊMEA	JOVEM
Antilopine wallaroo (*Macropus antilopinus*)	karndakidj kalaba	karndayh	djamunbuk
Black wallaroo (*Macropus bernardus*)	nadjinem baark	djukerre	——
Wallaroo (*Macropus robustus*)	kalkberd kanbulerri	wolerrk	narrobad
Agile wallaby (*Macropus agilis*)	warradjangkal	merlbbe	nakornborrh nanjid

O mais interessante, porém, é que para descrever o modo de saltar de cada um desses animais, a língua dispõe de um verbo específico:

NOME INGLÊS	VERBO 'SALTAR' EM KUNWINJKU	
	MACHO	FÊMEA
Antilopine wallaroo	kamawudme	kadjalwahme
Wallaroo	kamurlbardme	
Agile wallaby	kalurlhurlme	

Assim, se você perguntar a um falante de kunwinjku como se diz 'pular' em sua língua, ele só poderá responder se você especificar de que animal está falando... Como esses animais são muito importantes para essa população, e como são muito parecidos em seu aspecto visual, a lexicalização de seus modos de saltar oferece uma identificação muito mais rápida e precisa.

Quando os primeiros linguistas ocidentais chegaram à ilha da Nova Guiné, logo acima da Austrália, fizeram uma descoberta surpreendente: a ilha era o lugar com a maior diversidade linguística do mundo — mais de 800 línguas divididas em aproximadamente 16 famílias distintas. Por causa das montanhas extremamente altas da ilha, as populações foram se isolando ao longo dos milênios e desenvolvendo suas próprias línguas. Outra coisa que chamou a atenção dos ocidentais era o estágio de civilização extremamente recuado da grande maioria daquelas populações, algumas das quais vivendo praticamente na Idade da Pedra ainda no século XIX. Ora, essas mesmas populações que levavam uma vida ainda próxima da pré-história falavam línguas extremamente complexas do ponto de vista morfossintático. Como era possível?

A resposta é muito simples: não podemos confundir desenvolvimento tecnológico com desenvolvimento linguístico. Os seres humanos — seja qual for o estágio de sua cultura — pertencem todos a uma mesma espécie — *Homo sapiens* — e, portanto, dispõem todos do mesmo cérebro com as mesmíssimas capacidades cognitivas. Uma criança nascida no interior do Piauí, se for levada para viver desde pequena no Vietnã, aprenderá a falar vietnamita como qualquer criancinha nascida lá, e vice-versa. Como o cérebro humano é o mesmo e tem de dar conta de simbolizar o mesmo mundo e as mesmas experiências de vida, as línguas humanas, por mais diferentes que sejam — e são —, dispõem de recursos para desempenhar lindamente essas tarefas. Já na Idade Média, o filósofo inglês Roger Bacon (1214-1294) escreveu: *Grammatica una et eadem est secundum substantiam in omnibus linguis, licet accidentaliter varietur* ("A gramática é uma e a mesma, no que se refere à sua essência, em todas as línguas, embora varie acidentalmente"). A relação entre mente e mundo é que faz surgir os *universais linguísticos* aos quais já nos referimos.

Não estamos aqui, no entanto, é bom logo avisar, assumindo a hipótese inatista, defendida por Chomsky [▶51], segundo a qual o ser humano já nasce geneticamente condicionado para falar línguas, como se os universais linguísticos estivessem codificados no nosso DNA. Nossa postura se aproxima bem mais da de Michael Tomasello (já citado em [▶52]), quando ele escreve:

> O lugar onde as necessidades intelectuais se encontram com os recursos naturais mais diretamente está, evidentemente, na ontogenia humana. De fato, a sociogênese e a história cultural podem ser vistas como uma série de ontogenias em que os membros tanto maduros quanto imaturos de uma cultura aprendem a agir efetivamente na medida em que são expostos a problemas e equipados com recursos, incluindo interações sociais com habilidosos solucionadores de problemas. As habilidades cognitivas mais básicas exigidas para a aquisição da língua e o aprendizado da matemática complexa, como dois exemplos especialmente interessantes, estão universalmente disponíveis aos seres humanos. No entanto, as muitas e várias estruturas desses dois artefatos culturais tal como manifestos nas muitas e diferentes sociedades humanas do mundo não estão nem são — e de fato não podem estar nem ser — diretamente codificadas nos genes à frente do tempo. O modelo geral, portanto, é o de que os seres humanos têm habilidades cognitivas que resultam da herança biológica em ação no tempo filogenético; eles usam essas habilidades para explorar recursos culturais que evoluíram ao longo do tempo histórico; e fazem isso durante o tempo ontogenético. (1999: 47-48)

As diferenças entre as línguas não impedem, por exemplo, a prática da *tradução*, talvez uma das mais antigas atividades desempenhadas pelos seres humanos. A prova cabal disso está precisamente nos exemplos que demos acima. Não temos em PB um verbo específico para designar o modo de saltar de um *wallaroo* (até porque não existem *wallaroos* no Brasil), mas se tivermos de traduzir *kamawudme* para nossa língua diremos: "Saltar como um *wallaroo* antilopino macho", assim como traduziremos o tuiuca *diiga apé-'yigi* por "me disseram que ele jogou". Para decidir que tradução usar no momento de verter para o PB a forma verbal do inglês *you could* vamos ter de analisar todo o contexto e o cotexto para usar a forma verbal mais adequada, de modo a preservar a coesão e a coerência do texto.

O grande linguista russo Roman Jakobson (1896-1982) escreveu essas palavras lapidares: "As línguas se distinguem muito mais pelo que elas têm que dizer do que pelo que elas podem dizer". O que significa isso? Que numa língua humana se *pode dizer* tudo o que se diz numa outra língua humana, mas que esse *poder dizer* está sujeito à *gramática* de cada língua, que obriga o falante a se expressar inevitavelmente segundo as regras de seu idioma. Em PB, por exemplo, o gênero gramatical está inscrito em todos os substantivos da língua: no exato momento em que escolhemos um substantivo, ele já vem com seu gênero gramatical, que nos obriga a fazer a concordância de gênero com os demais termos do sintagma. Por isso, ao traduzir o inglês "*this house is beautiful but it is very expensive*", somos obrigados a dizer: *Essa casa é bonita mas é muito cara*. Se formos traduzir para o francês a célebre declaração de Júlio César — *Veni, vidi, vici* (*Vim, vi, venci*) —, seremos obrigados a usar três vezes o índice de pessoa *je* — *Je vins, je vis, je vainquis* —, porque em francês nenhum verbo pode vir desacompanhado de seu sujeito...

O empreendimento já antigo de versão da Bíblia para muitas línguas — os dois Testamentos para mais de 3.200 línguas, e um dos dois Testamentos para mais de 1.700 línguas —, censurável sob diversos aspectos (sendo o principal deles a dominação

espiritual dos povos colonizados para submetê-los mais facilmente à escravidão e à exploração), comprova ainda assim a possibilidade inequívoca de intercomunicação das sociedades humanas por meio de suas línguas específicas.

2.3 A DISCRIMINAÇÃO E O PRECONCEITO LINGUÍSTICOS

Uma derivação do mito da língua "primitiva" é a ideia de que as pessoas que não têm educação formal e não se valem das formas linguísticas padronizadas e prescritas pela tradição gramatical falam "tudo errado". Assim como os europeus se consideravam superiores aos negros, aos índios, aos polinésios, aos aborígines australianos etc., também muitas pessoas das camadas dominantes da sociedade consideram que os pobres, os analfabetos, os habitantes da zona rural (e, em alguns lugares, as mulheres, os jovens, os negros, os judeus, os imigrantes etc.) não sabem falar, têm vocabulário pobre e são incapazes de raciocínio lógico. É a expressão mais clara e vigorosa do preconceito linguístico, conjunto de ideias que se manifesta concretamente na discriminação pela linguagem.

Como já repeti várias vezes, esse preconceito foi e (infelizmente) continua sendo transmitido e preservado pela pedagogia tradicional de língua. Muitas e muitas pessoas abandonam os estudos porque ficam traumatizadas ao entrar na escola e, logo em seus primeiros contatos com o mundo escolar, ser alvo de discriminação, de deboche, de piada por causa de seu jeito de falar — discriminação praticada não só pelos colegas, mas também por muitas professoras e muitos professores.

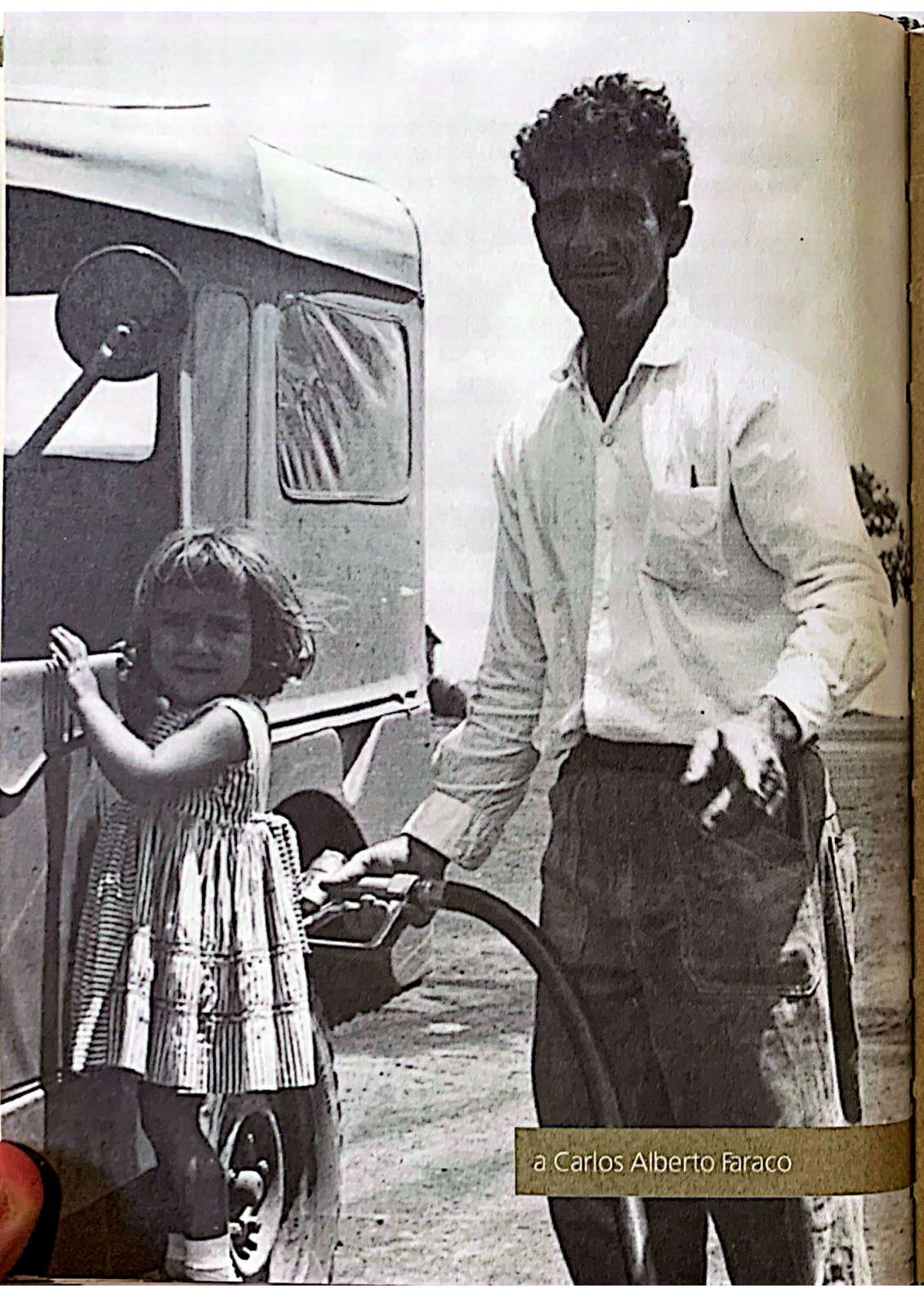
Vai nessa mesma linha de pensamento arcaico uma coluna assinada pelo poeta Ferreira Gullar, na *Folha de S.Paulo* (17/2/2008), em que ele (logo ele, que é nordestino e que sempre foi considerado um homem de ideias progressistas!) pergunta:

> (2) Mas quem é o povo? Aquela gente nordestina, magricela, tostada de sol, *que mal sabe falar?* (grifo meu)

A afirmação de que existe alguém que "mal sabe falar" revela, para começar, uma profunda ignorância de seu enunciador no que diz respeito aos avanços da antropologia, da psicologia, da medicina, da linguística contemporâneas etc. Todo e qualquer ser humano dotado de suas plenas capacidades físicas e mentais é perfeitamente capaz de falar e de *falar bem*. O cérebro humano é extremamente organizado, de modo que, ao contrário do que tanta gente pensa, é simplesmente impossível falar sem obedecer regras gramaticais. O problema está em acreditar que as únicas regras gramaticais são as *poucas, pouquíssimas*, descritas (mal descritas) e prescritas pela tradição gramatical. Se alguém falasse "sem gramática" não conseguiria se fazer entender

— e nem a pessoa mais preconceituosa vai poder dizer que um enunciado como *"nós chegou lá e os home já tinha ido embora"* é incompreensível porque, supostamente, não apresenta concordância nem obedece estritas regras morfossintáticas...

Não existem "línguas primitivas". Existem, isso sim, opiniões, preconceitos, mitos, lendas, ilusões, fantasias — elas, sim — primitivas, toscas, tacanhas e bisonhas – quando não explicitamente fascistas. E é desse tipo de ignorância que temos obrigação de nos livrar e de livrar aqueles que nos procuram como educadores.



a Carlos Alberto Faraco

3

de língua materna a língua paterna

do vernáculo à normatização

O linguista Bernard Cerquiglini, em dois de seus livros sobre a história do francês (1991, 2007), emprega a expressão língua paterna para se referir ao produto das políticas de *normatização* que incidem sobre as línguas de sociedades grafocêntricas, como as europeias. Evidentemente, *língua paterna*

se contrapõe a língua materna, um conceito muito difundido e de uso amplo dentro e fora das discussões especializadas. Cerquiglini emprega muito esporadicamente a expressão nos livros citados, mas vale a pena desenvolvê-la um pouco mais aqui para as discussões que vamos fazer nesse capítulo.

A língua materna é precisamente a língua da mãe, a língua que cada pessoa começa a adquirir tão logo nasce e cria o vínculo afetivo-linguístico com a mãe (ou, na falta dela, com a pessoa que venha a preencher esse papel). É uma língua puramente *oral* — falada e ouvida —, mesmo quando provém da voz de uma pessoa altamente letrada. Língua do afeto, do desejo, do íntimo, do sonho, vive à margem dos ditames da *norma* canonizada. A língua materna é intrinsecamente variável, doméstica, familiar, idioma particular daquilo que em inglês se chama *household*, um termo que inclui a casa, seus ocupantes e todas as atividades ali desenvolvidas por eles. Quando, na história de cada sociedade, uma determinada língua — ou mais precisamente, uma das *variedades* dessa língua — é alçada à condição de língua *oficial*, ocorre uma importante metamorfose — a língua materna se torna *língua paterna*, transformada em *padrão* (do latim *patronu-*, onde está presente a raiz *pater-* 'pai', mesmo vocábulo do qual procedem *patrão* e *patrono*). É a língua patrocinada pelo Estado e, irradiando-se dele, a língua da escola — isso explica o choque que muitas pessoas (especialmente os falantes de variedades linguísticas estigmatizadas) experimentam ao tomar contato pela primeira vez com uma língua que, em boa medida, além de estranha é quase estrangeira. A língua paterna é essencialmente *escrita*, *ortografizada*, *normatizada*.

A língua paterna é a língua da Lei, sempre associada à figura do pai, inclusive nos postulados da psicanálise freudiana. A língua materna — língua de *mulher* — sofre na maioria das sociedades as mesmas depreciações dedicadas ao gênero feminino: é o lugar do "erro", do "desvio", do "frágil", do pouco confiável, do instável, do inconvenientemente sensível e sensitivo. Ao pai cabe domar e domesticar esse idioma errático, conferindo-lhe regras, regimentos, registros, regências, regulamentos — palavras todas derivadas de *rex*, *regis*, 'rei', assim como *recto*, *direcção*, *correcção*, *régua*. É a língua do direito (< *derectu-*, 'o que está reto'), erigida como lei linguística. A língua paterna é a língua da erecção, a língua do rei, pai da nação, símbolo do Estado. (Sobre essa raiz e seus derivados, ver [▶283].).

Um elemento fundamental nessa padronização é, sem dúvida alguma, a presença da *escrita*, tomada sempre em sua vertente *canônica*, *douta*, *literária*. Tudo o que tem a ver com a instituição de uma linguagem "certa", "oficial", "uniforme", "normatizada" etc. também tem a ver com o uso intenso da *escrita*. Daí advém o uso tradicional da expressão *norma culta*, identificada sempre com a linguagem escrita mais formal, mais monitorada, de preferência com pretensões "literárias". As sociedades que são fortemente *letradas*, isto é, em que a *cultura escrita* é onipresente e supervalorizada, são também aquelas que ostentam instituições com grande poder

centrípeto sobre as forças de mudança da língua [▶125]. O linguista canadense J. M. Paquette explica de que maneira a normatização da língua se associou estreitamente à normatização jurídica durante o período da história europeia em que os Estados nacionais se fortaleceram e se sentiu a necessidade de todo um corpo de instituições e de funcionários capazes de elaborar normas, regulamentos e leis, processo que exigiu, simultaneamente, a uniformização das ortografias e a padronização das regras gramaticais:

> Tanto quanto possamos apreender suas origens na história, o serviço da chancelaria real já pode ser percebido como um universo da *escrita* estreitamente ligado à atividade *jurídica*. Não há mais dúvidas, a seguir, de que, na história das diversas chancelarias da Europa, assim será até nas épocas mais recentes — mas é interessante observar que desde sua emergência na história das instituições, o conjunto das funções da chancelaria vinculam a *escrita* e o *direito*. (Paquette, 2001: 244)

Não é por mera coincidência, então, que também neste período (séculos XIV-XVI) tenham sido escritas as primeiras gramáticas normativas das línguas europeias. Esse vínculo estreito entre *escrita* e *direito* é o que explica por que as gramáticas normativas "se apresentam até os nossos dias sob a forma de um verdadeiro código de direito, com a regra, os parágrafos, os artigos, as exceções quanto aos exemplos tirados dos autores" — segundo Paquette, "é porque elas têm mais ou menos uma função análoga à da jurisprudência" (p. 246).

A masculinidade da língua padronizada é explicitada no célebre introito da gramática de João de Barros (1540), em que o autor escreve que os antigos definiram a palavra gramática como "hũ módo çérto e iusto de falár, & escreuer, colheito do uso, e autoridáde dos barões doutos". "Barões" é outra forma de "varões", ou seja, *homens*, pois são eles — e jamais as mulheres — os conhecedores e praticantes do "modo certo e justo de falar e escrever".

Uma vez instituída *fora do lar*, na esfera pública, a língua paterna logo encontra seus defensores, eminentemente masculinos, capazes de preservá-la em sua *pureza*. É o trabalho sempre trágico e desesperado dos *puristas*. Nas palavras de Cerquiglini (2007: 25),

> atentar [contra a língua paterna], por meio de algumas reformas (reajustes ortográficos, feminização dos nomes das profissões), ou pela tolerância para com as palavras estrangeiras não é somente ofender o uso, é desatar os fios do eu. A novidade induzida dispara uma reação cuja regularidade quase previsível, cuja sinceridade decerto, cuja violência por vezes, mostram que a correção gramatical não é só o que está em causa. Inversamente, amar e defender a norma, mesmo incoerente, vale por atestado de cidadania, penhor de adesão à comunidade nacional [...]

Enquanto escrevo (janeiro de 2011), acaba de assumir a presidência da República a primeira mulher brasileira a ocupar esse cargo: Dilma Rousseff. A grande mídia, sempre vinculada aos interesses das minorias dominantes, se recusa a chamá-la de presidenta, muito embora ela própria faça questão de deixar bem marcada a singularidade histórica de seu mandato, chamando a si mesma de *presidenta*, e muito

embora a palavra esteja devidamente registrada nas gramáticas normativas e nos dicionários. É a reação dos defensores da *língua paterna*, ainda atônitos com a visão de uma mulher no cargo mais alto do país.

A língua paterna se ergue como patrimônio a ser preservado. Vem devidamente codificada nas gramáticas e catalogada nos dicionários. É a *norma-padrão*, identificada com a Pátria. A língua materna vive à margem da Lei, no interior das casas, não deve sair do gineceu. No entanto, é nesse mundo desprezado (ou alvo de paternalismo condescendente) que pulsa a força vital da língua, seus processos de mudança, suas metáforas, sua interminável poética. É nele que vamos conhecer, não mais a língua do Estado, porém o real estado da língua: o *vernáculo*.

3.1 O VERNÁCULO NA SOCIOLINGUÍSTICA

Quando consultamos um dicionário de referência, como o *Houaiss*, encontramos no verbete vernáculo as seguintes acepções:

■ adjetivo
1 próprio de um país, nação, região
Ex.: <língua v.> <costumes v.>
2 Derivação: sentido figurado.
diz-se de linguagem correta, sem estrangeirismos na pronúncia, vocabulário ou construções sintáticas; castiço
■ substantivo masculino
3 a língua própria de um país ou de uma região; língua nacional, idioma vernáculo
Ex.: o ensino do v.

O termo vernáculo tem sua origem no latim, *vernaculu-*, 'relativo ao escravo nascido na casa do amo' e, por extensão de sentido, 'de casa, nascido ou produzido no país'. O mais curioso é que o próprio termo não era "vernáculo" em latim, já que deriva de *verna*, 'escravo nascido na casa do amo', uma palavra que o latim tomou de empréstimo ao etrusco.

As acepções listadas pelo *Houaiss* são todas clássicas e figuram nos dicionários de muitas línguas europeias. No entanto, como termo técnico da sociolinguística, *vernáculo* não corresponde a nenhuma daquelas definições.

O americano William Labov, considerado o fundador da sociolinguística variacionista, queria romper com a metodologia de pesquisa que vigorava na linguística americana dos anos 1960, uma metodologia indutiva, em que o estudioso recorria à sua própria competência de falante nativo para, com base nela, abstrair as regras de funcionamento da gramática da língua — é a metodologia característica do gerativismo fundado por Noam Chomsky [▸47].

Labov queria conhecer a língua em sua realidade social para capturar o que mais lhe interessava: a variação. Ao invés de descartar os usos variáveis como epifenômenos, ou seja, como ocorrências acidentais, acessórias, que não alteravam a configuração do sistema da língua, Labov queria demonstrar que a variação era constitutiva do próprio sistema e, principalmente, que era o motor social da mudança das línguas. Para isso, era preciso recolher amostras da língua viva, da língua em uso. Com esse objetivo, ele elaborou a metodologia da entrevista sociolinguística, que só foi possível graças à invenção do gravador portátil que, nos anos 1960, ainda usava os grandes rolos de fita magnética que lhe valeram seu nome no Brasil: *gravador de rolo*. Só mais tarde ia surgir a fita *cassette* (palavra francesa que significa 'caixinha') e, mais recentemente ainda, a gravação em formato digital, que reduziu drasticamente o tamanho dos aparelhos. A forma e o tamanho do aparelho são dados importantes para compreendermos as dificuldades metodológicas que Labov enfrentou e as táticas que elaborou para contorná-las.

> Para Labov, os dados mais preciosos para a investigação sociolinguística estavam na língua falada mais espontânea, menos monitorada, mais livre de coerções sociais e culturais, empregada em situações distensas, descontraídas. Essa língua falada maximamente espontânea e não monitorada é o que ele passou a designar com o termo inglês vernacular, que traduzimos por vernáculo.

Ora, como gravar essa fala espontânea se a mera presença do pesquisador, um indivíduo de fora da família ou do grupo, já criava uma situação tensa e, para piorar, se ele trazia consigo um aparelho enorme com um microfone do tamanho de uma lata de cerveja? Essa situação foi rotulada por Labov de paradoxo do observador: querer registrar a fala mais espontânea num contexto de interação verbal não espontâneo, como é a entrevista sociolinguística.

Uma das soluções engenhosas de Labov foi deixar a pessoa entrevistada falar à vontade até que, em dado momento, perguntava a ela: "Você já passou por alguma situação em que achou que ia morrer? Uma situação em que você disse a si mesma: *chegou a minha vez*?" Diante dessa pergunta, a pessoa entrevistada passava a narrar alguma experiência perigosa em sua vida e, levada pela emoção da narrativa, se preocupava mais com *o que* estava dizendo e menos com o *como*, deixava de monitorar sua fala e fazia emergir o seu *vernáculo*, ou seja, o seu modo mais natural e descontraído de falar. Assim Labov conseguia os dados que tanto desejava. Outra boa fonte de dados do vernáculo eram os momentos em que, durante a entrevista, a pessoa desviava sua fala do entrevistador e a dirigia a alguém da família, como uma filha, por exemplo, e lhe dizia algo de natureza íntima, em linguagem própria do ambiente doméstico. Mais tarde, Labov passou a usar a tática de recrutar alguém de dentro da própria comunidade para coletar os dados. Isso foi fundamental para seu famoso trabalho a respeito do vernáculo dos negros americanos que viviam em guetos marginalizados nas grandes cidades.

O vernáculo, portanto, como termo técnico da sociolinguística, tem essas características: é falado, é espontâneo, é informal, é não monitorado. Ele é o foco dos estudos linguísticos porque é nessa variedade que a língua se apresenta tal como ela realmente é em seu estado atual, livre das pressões da língua escrita e dos padrões ditados pela norma gramatical imposta pelas instituições sociais como a escola (as forças centrípetas que vamos estudar no próximo capítulo). É claro que o não monitoramento absoluto é incapturável, pois o simples ato de falar já condiciona os usos que a pessoa faz dos recursos linguísticos disponíveis em sua língua. Mesmo assim, o vernáculo obtido na coleta sociolinguística tem muita coisa a nos revelar sobre a língua atual. É no vernáculo que se manifestam as tendências da língua na direção da mudança, que será o objeto do nosso próximo capítulo.

3.2 O VERNÁCULO BRASILEIRO

A metodologia desenvolvida por W. Labov tem sido empregada no Brasil há cerca de quarenta anos pelos linguistas interessados em obter o retrato mais fiel possível do nosso vernáculo. Todo esse volumoso banco de dados de língua falada vem servindo para a descrição da nossa língua em suas diversas variedades regionais, sociais, etárias, étnicas, de gênero etc. Graças aos resultados dessas pesquisas, já não é mais possível dizer, como há meio século, que o PB ainda é pouco conhecido e pouco descrito.

As gramáticas normativas sempre se basearam nos usos escritos, literários, para fazer o seu trabalho de prescrição. Por isso, é claro que nelas não vamos encontrar uma descrição fiel e detalhada do que é, de fato, a língua em seus usos mais espontâneos, característicos dos estilos falados menos monitorados, nos dias de hoje. Também nelas quase nunca aparecem as formas inovadoras que já se firmaram também na escrita (ou, quando aparecem, é para ser alvo de censura). As gramáticas prescritivas, em geral, fazem uma apresentação anacrônica da língua, isto é, desvinculada dos usos reais contemporâneos, até mesmo da escrita literária atual. Por isso, as gramáticas prescritivas não podem servir de material pedagógico para quem ensina língua hoje, no Brasil, ou para quem ensina o PB como língua estrangeira.

Se é justificável propor um modelo de língua para ser ensinado nas escolas e para servir de referência nas situações de maior monitoramento da fala e da escrita, é imprescindível que esse modelo seja baseado na realidade da nossa língua, nas regras que de fato a governam, e não em idiossincrasias e veleidades de gramáticos, dicionaristas e de escritores, ainda mais se forem portugueses, um povo que não fala a mesma língua que nós. O objetivo dessa nossa gramática é justamente contribuir para que os docentes conheçam melhor a sua própria língua, se reconheçam nela e se sirvam dela como seu objeto de trabalho na educação linguística de seus aprendizes.

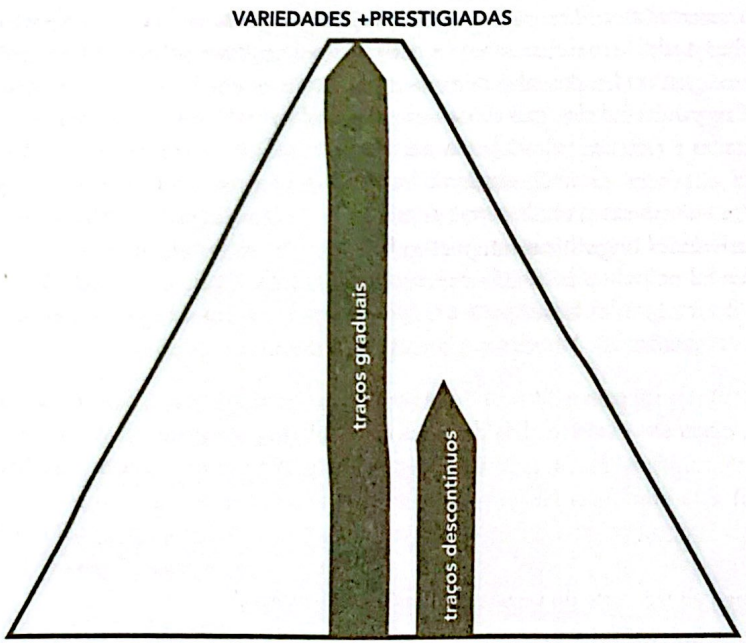
Pelas características da sociedade brasileira, extremamente excludente e injusta na distribuição dos bens sociais, sabemos que os traços linguísticos próprios das variedades mais estigmatizadas, dos falantes rurais ou rurbanos — isto é, que vivem nas periferias das grandes cidades, mas ainda conservam hábitos culturais da zona rural —, são rejeitados e evitados pelos falantes das outras variedades. Com isso, muitos desses traços *não fazem parte* do vernáculo brasileiro mais geral. Eles configuram o que muitos autores chamam de *norma popular* e que nós aqui temos preferido chamar de variedades linguísticas estigmatizadas (VLE). Por outro lado, os traços linguísticos usados pelos brasileiros urbanos escolarizados em suas interações faladas/escritas menos monitoradas fazem parte *também* do repertório dos falantes das variedades mais estigmatizadas. Vamos ver o que isso quer dizer.

A pronúncia [o] para o ditongo escrito *ou*, como em *roupa, pouco, ouro, louco, comprou, amou* etc. ocorre na fala de **todos os brasileiros**, independente de sua origem geográfica, classe social, grau de letramento etc. No que diz respeito à **avaliação social**, essa pronúncia não provoca nenhuma reação negativa: nem a pessoa mais letrada interrompe uma conversa para corrigir seu interlocutor ou censurá-lo por ter pronunciado "rôpa", "pôco", "ôro" etc. Conclusão: **a pronúncia [o] para o antigo ditongo *ou* faz parte do vernáculo brasileiro mais geral.**

A pronúncia [ɪ] para o dígrafo escrito *lh*, como em *telha, abelha, palha, trabalha* caracteriza **os falantes das variedades mais estigmatizadas, sem prestígio social.** No tocante à avaliação social, essa pronúncia provoca **reação negativa** por parte dos falantes urbanos escolarizados, que a identificam imediatamente como próprias de pessoas "pobres", "sem instrução", "caipiras" etc. Conclusão: **a pronúncia [ɪ] para o dígrafo escrito *lh* não faz parte do vernáculo brasileiro mais geral.**

Temos assim, claramente, dois grandes conjuntos de traços linguísticos: aqueles que aparecem na fala de *todos os brasileiros*, independente de sua origem social, regional etc.; e aqueles que aparecem primordialmente na fala dos brasileiros de origem social humilde, de pouca ou nenhuma escolaridade, de antecedentes rurais etc. Ao primeiro grupo de traços os sociolinguistas dão o nome de **traços graduais.** Ao segundo, o nome de **traços descontínuos.** Essa terminologia foi proposta pela sociolinguista brasileira Stella Maris Bortoni-Ricardo para traduzir os termos *gradient* e *sharp* empregados pelos pesquisadores de língua inglesa.

Embora não seja função do ensino estimular essa substituição, somos forçados a reconhecer que ela ocorre e que parte dos próprios falantes. É claro que também existem aqueles que preferem se manter fiéis à sua herança linguística comunitária, até como forma de protesto contra a imposição de uma norma que é estranha à sua identidade. O que a escola não pode deixar de fazer, como venho insistindo, é ensinar aquilo que os falantes não sabem, não trazem de sua vivência familiar e comunitária. O destino que eles vão dar ao que aprenderem não cabe à escola determinar.



Por sua vez, os traços *graduais* recebem esse nome porque sua ocorrência nas variedades mais prestigiadas é função primordialmente do **grau de frequência** com que eles são usados pelos falantes. Um exemplo clássico é o da concordância verbal e nominal: **nenhum falante de PB aplica integralmente as regras de concordância** previstas pela tradição normativa; no entanto, os falantes urbanos mais escolarizados tendem a fazer mais uso dessas regras, principalmente em seus estilos mais monitorados de fala e de escrita, do que os falantes menos escolarizados.

Os *traços descontínuos* são precisamente aqueles aspectos linguísticos que sofrem **a maior carga de discriminação e preconceito** na nossa sociedade, muito embora (senão por isso mesmo) pertençam à grande maioria da nossa população. Por caracterizarem as variedades linguísticas de falantes com baixo ou nenhum prestígio social, esses usos linguísticos são rejeitados, repelidos, ridicularizados e evitados tanto quanto possível pelos cidadãos que se acham (ilusoriamente) portadores da língua "certa".

Essas características das variedades estigmatizadas vêm sendo investigadas e descritas por muitos pesquisadores. Elas têm uma lógica linguística perfeitamente demonstrável, não surgiram por causa da "ignorância" nem da "preguiça" de seus usuários e representam, algumas vezes, como vamos examinar no próximo capítulo, sobrevivências de fases anteriores da língua e/ou sinais de mudança. No entanto, por razões de ordem exclusivamente social e cultural, essas características da fala "popular" são vistas como

a forma mais "errada" possível de falar a língua. Algumas dessas características mais marcantes são as seguintes (para saber mais sobre elas, consulte a pagina indicada):

1. Queda da vogal átona postônica em palavras proparoxítonas: *córrego* > *corgo*; *pássaro* > *passo*; *bêbado* > *bebo*; *árvore* > *arvre* etc. [▶316].
2. Não nasalização de sílabas postônicas: *homem* ~ *home*; *ontem* ~ *onte*; *fizeram* ~ *fizero* etc. [▶329] .
3. Monotongação de ditongos átonos crescentes em posição final: *notícia* > *notiça*; *paciência* > *paciença*; *imundície* > *imundice* / *mundiça* etc. A história da língua apresenta muitos exemplos dessa tendência, que os falantes das variedades estigmatizadas seguem mais coerentemente: *iustitia* > *justiça*; *pigritia* > *preguiça*; *factitiu* > *feitiço*; *pretiu* > *preço*; *criantia* > *criança* etc. A terminação *-ícia*, mais próxima do étimo latino, é resultado do processo de relatinização [▶143].
4. Rotacismo: troca de [l] por [r] em encontros consonantais ou em final de sílaba: *placa* > *praca*; *planta* > *pranta*; *talco* > *tarco*; *futebol* > *futebor* etc. [▶156].
5. Pronúncia [ɪ] da consoante palatal [ʎ], escrita *lh*: *telha* > *teia*; *abelha* > *abeia*; *velha* > *veia* etc. [▶393].
6. Redução da terminação *-ndo* do gerúndio em *-no*: *cantano*, *comeno*, *dormino* etc. Um caso de *assimilação* [▶150].
7. Eliminação do plural redundante, marcado em geral só nos determinantes: *os menino*, *as casa*, *aquelas coisa toda* etc. Um caso de *economia linguística* [▶147].
8. Redução da morfologia verbal a duas formas: [eu] *canto* e [*tu/você/ele/nós/a gente/vocês/eles*] *canta*; ou a três: [*eu*] *canto*; [*tu/você/ele/vocês/eles*] *canta*; [*nós*] *cantamo*. Outro caso de economia linguística [▶147].
9. Léxico característico, variável de região para região: *fruita*, *luita*, *oitubro*, *cunzinha*, *drumi*, *percurá*, *despois*, *antonce*, *arrespondê*, *alembrar*, *dereito*, *menhã*, *sumana*, *vosmecê* etc. A maioria desses vocábulos representam sobrevivências de fases anteriores da língua e podem ser encontrados na literatura medieval e clássica.

Essas e outras características linguísticas pertencem ao *vernáculo* dos falantes das variedades mais estigmatizadas que compõem, como já dissemos, a ampla maioria da nossa população. Para que deixem de ser traços descontínuos e consigam derrubar as barreiras sociais, é necessário que se dê alguma transformação muito radical na sociedade brasileira, ou que, pela ampliação da mobilidade social, os falantes dessas variedades se integrem cada vez mais às camadas médias e médias altas. A história das línguas e das sociedades revela, no entanto, que a mobilidade social leva, na grande maioria dos casos, à substituição dialetal, ou seja, como já mencionamos, ao abandono das características próprias dos falantes estigmatizados em favor dos tra-

ços graduais ou mesmo de usos peculiares às camadas sociais mais prestigiadas. Os estudos de Labov mostram que os falantes das classes médias baixas e baixas são os que produzem com maior frequência as pronúncias consideradas de maior prestígio, mais até do que os falantes das médias altas e altas. Isso se deve ao que ele chama de insegurança linguística: o sentimento de que sua variedade não é boa e que é preciso enfatizar os traços de prestígio para que esse prestígio se transfira para a própria pessoa. Também muito frequente nessas classes médias baixas e baixas é o fenômeno da hipercorreção (ver o capítulo 21), que leva os indivíduos a generalizar uma regra normatizada, mesmo onde ela não se aplica. Exemplo típico é a flexão no plural do verbo *haver* com sentido apresentacional: "Houv*eram* muitos problemas no embarque", ou a flexão de gênero no advérbio *menos*, interpretado como adjetivo: "Tinha *menas* gente lá do que eu esperava".

3.3 A LÍNGUA QUE ESSA GRAMÁTICA DESCREVE

Esse livro tem a intenção de fazer uma descrição das principais características do vernáculo geral brasileiro (VGB), ou seja, dos usos que ocorrem *em todas as variedades* — o que acima foi designado como *traços graduais*. Embora esses usos ainda sejam injustificadamente (e irracionalmente) condenados pelas instâncias normativas e puristas da nossa sociedade, criando uma situação extremamente dramática no ambiente educacional — afinal, o que temos de ensinar? —, eles são a espinha dorsal do português brasileiro, são eles que configuram uma língua que é a terceira mais falada do Ocidente, língua de uma nação que desempenha um papel cada vez mais proeminente na geopolítica e na economia mundiais.

Esses usos caracteristicamente brasileiros já estão plenamente integrados à gramática da nossa língua, gramática entendida aqui como o conhecimento intuitivo que todo falante nativo tem de seu próprio idioma. Eles resultam de mais de meio milênio de mudanças linguísticas (capítulo 5) provocadas pelo contato do português europeu quinhentista com outros povos, falantes de outras línguas, principalmente africanas, e dos fatores cognitivos que suscitam essas mudanças em todas as línguas do mundo.

Como venho repetindo, a função da escola é ensinar aquilo que o aprendiz não conhece. Por isso, a educação linguística deve promover a aquisição pelos futuros cidadãos das formas de falar e de escrever que eles ainda não dominam. A escola não precisa ensinar ninguém a dizer *para mim fazer*, porque isso já faz parte da língua materna de milhões de brasileiros: precisa, isso sim, ensinar a dizer e a escrever *para eu fazer*, de modo que o aprendiz se apodere *também* dessa construção sintática para fazer uso dela, se quiser ou se achar conveniente.

Por outro lado, a escola também não precisa ensinar as regras e sub-regras da colocação pronominal clássica, porque elas não fazem parte da gramática do PB, onde só vigora uma única regra de colocação dos clíticos: a próclise ao verbo principal [▸762]. Os demais usos podem ser discutidos a partir da percepção deles em textos autênticos lidos e trabalhados em sala de aula, mas não para serem *ensinados* como as únicas formas certas e boas de falar ou de escrever.

Essa nossa gramática quer contribuir na criação de um modelo de referência para as práticas sociais de linguagem que exigem maior grau de monitoramento estilístico na fala e na escrita. Pode ser considerada normativa nesse sentido, mas não *prescritiva*, porque os usos descritos aqui e propostos como modelo não necessitam de imposição nem de prescrição, já que pertencem à língua falada por todos os brasileiros.

O normativismo tradicional tem sido duramente criticado pelos linguistas e pelos educadores justamente por tentar impor, com a ajuda de instrumentos ideológicos repressivos, um modelo de língua em tudo irracional, a começar por seu apego à escrita literária de autores consagrados, de preferência antigos, e por tomar essa escrita literária como espelho em que todo e qualquer cidadão comum deve se mirar para falar ou escrever, independente do gênero textual e do contexto de interação verbal em que se encontre. Filiado a uma secular ideologia que despreza a língua falada e tenta sujeitar uma língua viva, moderna, aos moldes da gramática do latim clássico — uma língua que nunca foi realmente falada, pois também era um modelo literário —, esse normativismo é pernicioso, porque é antidemocrático, pré-científico, quando não francamente anticientífico, o que se comprova pelos arroubos obscurantistas que frequentemente invadem os meios de comunicação na forma de ataques à ciência linguística e a qualquer esforço de tornar mais respirável e menos sufocante o debate sobre língua e ensino de língua em nosso país.

Existe lugar, no entanto, para um normativismo linguístico comparável ao normativismo das leis que, numa sociedade verdadeiramente republicana e democrática, visam o bem-estar e a convivência pacífica dos cidadãos, um normativismo que se faça a favor da maioria das pessoas e não contra elas. Como escreveu o linguista francês Alain Rey:

> Não se trata de rejeitar toda norma — nenhuma sociedade pode abrir mão delas —, mas sim de fiscalizar sua construção com a análise científica, e compreender a atividade normativa, isto é, modificá-la, como um setor da prática social menos inocente do que parece. (2001: 144)

3.4 POR QUE *PORTUGUÊS BRASILEIRO*?

Diz a lenda que o escritor irlandês George Bernard Shaw, famoso por suas tiradas maliciosas, teria afirmado certa vez: "Estados Unidos e Grã-Bretanha são dois países

separados pela mesma língua". Por outro lado, em seu famoso e divertidíssimo conto "O fantasma de Canterville", o também irlandês Oscar Wilde escreveu: "*Nós [ingleses] realmente temos tudo em comum hoje em dia com os Estados Unidos, exceto, é claro, a língua*". Essas afirmações, levadas a sério, poderiam se aplicar perfeitamente ao Brasil e a Portugal.

Países nascidos do processo colonial até hoje usam o nome da língua de seu colonizador para designar o idioma mais empregado no território nacional. São mais de 300 milhões de americanos contra 60 milhões de britânicos, mas ainda assim a língua mais empregada nos Estados Unidos se chama *English*, 'inglês'. Na América Latina vivem mais de 400 milhões de falantes de distintas variedades de "espanhol", mas todas atendem pelo mesmo nome.

Quando a comparação se faz entre Brasil e Portugal, os números são ainda mais díspares: 190 milhões de brasileiros contra pouco mais de 10 milhões de portugueses (a população do município de São Paulo; a área metropolitana tem o dobro disso); uma área de mais de 8 milhões de quilômetros quadrados (o dobro da área total da União Europeia), contra 92 mil quilômetros quadrados, menor do que Santa Catarina (98.000km^2), que é o vigésimo em área entre os 27 estados brasileiros. Se passarmos aos números da economia e da importância geopolítica no cenário mundial, Portugal simplesmente desaparece (e desaparece, infelizmente, no fundo do poço em que está ao lado de outros países como Grécia e Irlanda, submetidos às imposições draconianas dos países mais ricos do continente para que pudessem integrar a União Europeia).

Nessa gramática, adotamos a terminologia português brasileiro para nos alinharmos a uma denominação que vem sendo amplamente utilizada na pesquisa científica. É importante perceber que existe uma diferença entre *português brasileiro* e *português do Brasil*, uma expressão que também já foi muito usada. A preferência por *português brasileiro* indica um reconhecimento da especificidade da nossa língua dentro de um grupo maior de línguas, que no capítulo 5 vou chamar de *portugalego*.

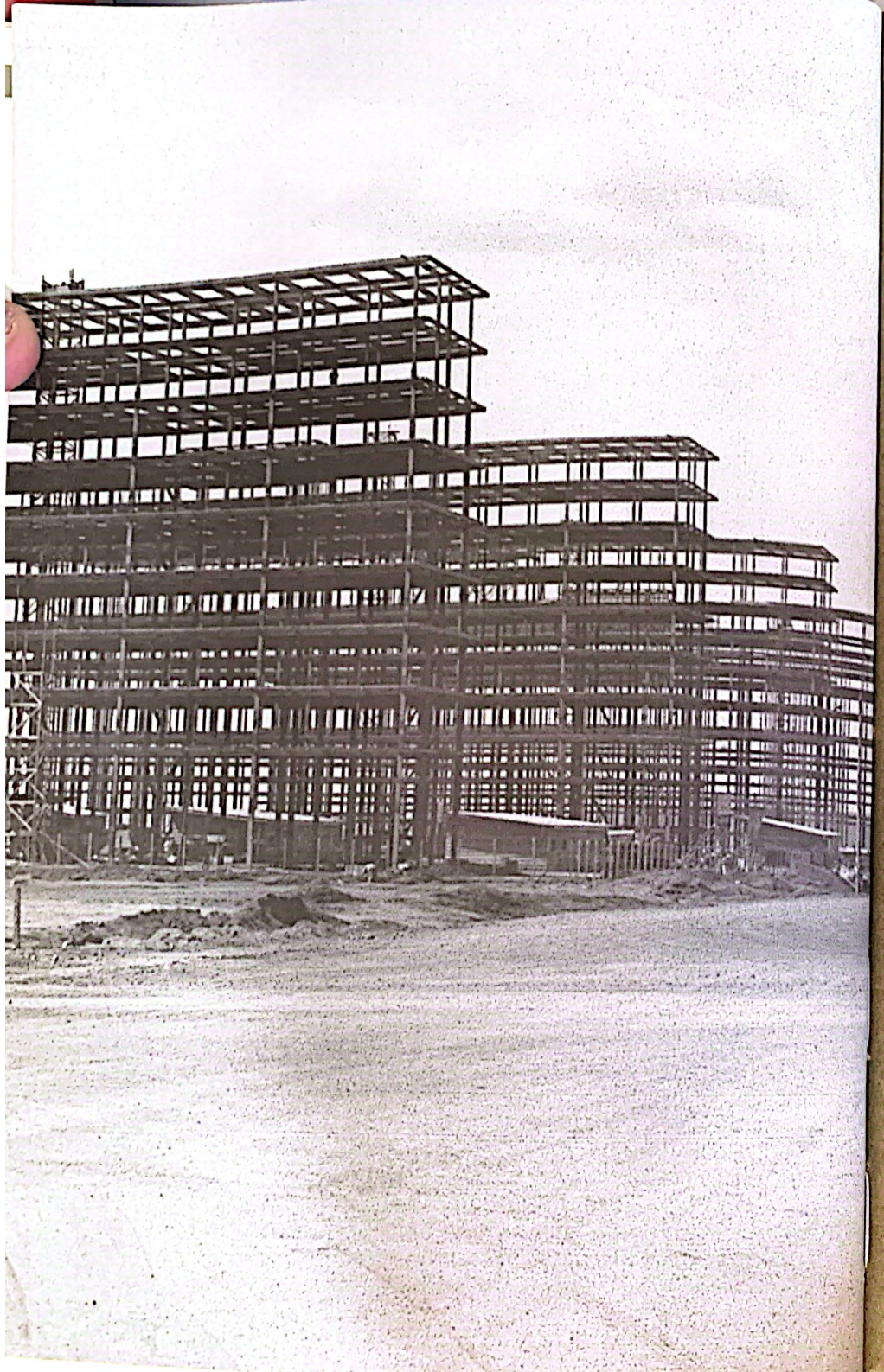
Em sua *Nova gramática do português brasileiro*, Ataliba de Castilho assim se expressa:

> Há mais de quinhentos anos a língua portuguesa foi trazida ao Brasil. Nos séculos XVI a XVIII foi rotulada como *o português no Brasil*, pois era inteiramente lusitana, e não tinha superado as línguas indígenas. A partir do século XIX, a língua portuguesa tornou-se majoritária, começou a distanciar-se do português europeu, sendo então denominada *português do Brasil*. A partir dos anos 80 do século XX, suprime-se a preposição *do*, e começamos a falar em *português brasileiro*. Sinaliza-se com isso que novos distanciamentos tinham ocorrido, servindo a expressão para designar a identidade linguística dos brasileiros. (Castilho, 2010: 31)

É certo que muitas pessoas gostariam que nossa língua já fosse chamada tranquilamente de *brasileiro*. Seja qual for o nome que se dê a ela, no entanto, o que considero mais importante do que qualquer outra coisa é o reconhecimento do PB como

uma língua plena, perfeita para atender a todas as necessidades de interação social e de construção da identidade dos cidadãos de uma nação soberana, rica e importante. Temos que estudar e ensinar a nossa língua com base no que ela é aqui e agora, no Brasil do século XXI.

As comparações com o português europeu, no domínio da educação linguística, devem ser deixadas de lado, seja as que tentam mostrar como a nossa língua é supostamente inferior à dos portugueses, seja as que tentam mostrar como é superior à deles. Portugal não pode ser o parâmetro contra o qual devemos medir a nossa língua. Portugal não dita os destinos da nossa língua, que é total e exclusivamente nossa, dos brasileiros. Não vamos negar nosso passado, nossa história, mas também não podemos persistir num espírito colonizado, quando falta tão pouco para o bicentenário da nossa independência. É nessa nova ideologia linguística que essa nossa gramática se inscreve, é essa a filosofia que nos orienta.



LIVRO II HISTÓRIA DO PORTUGUÊS BRASILEIRO

a Henrique Monteagudo

4

nada será como antes

a mudança linguística

4.1 Concepções da mudança linguística ao longo do tempo

4.1.1 Visões tradicionais da mudança

A certeza de que as línguas mudam com o tempo é facilmente obtida pelo estudo dos documentos escritos no passado. Se compararmos um texto publicado hoje no Brasil, por exemplo, com a carta escrita por Pero Vaz de Caminha ao rei de

Portugal em 1500 dando notícia do "achamento" da terra nova, as diferenças saltam aos olhos, por mais que a escrita, como se sabe, tenda a ser mais conservadora e não registre fielmente a realidade da língua viva, falada, de seu tempo.

A tranquilidade na aceitação da mudança linguística, no entanto, é bem mais difícil de ocorrer. As pessoas em geral se conformam com a ideia de que "o português veio do latim" (ideia, aliás, que vamos criticar em [▸221], já que o português veio mesmo é do galego, mas por enquanto não vamos tratar disso) ou que a língua da época de Camões era diferente da de hoje. No entanto, quando o foco se fecha sobre o que está acontecendo aqui e agora na língua, sobre as mudanças inevitáveis que estão se processando no PB, a reação social e cultural é muito diferente. O falante comum não tem uma visão em perspectiva nem em prospectiva, de longo prazo, da história de sua língua. Não se detém para pensar que, se a língua mudou no passado, não tem por que não mudar também agora, no presente, antecipando o que ela vai ser no futuro. O futuro da língua, na imaginação da maioria dos falantes, é um retrato fiel do que a língua é agora. Conseguem visualizar em suas mentes naves espaciais velocíssimas, prédios de vidro transparente com oitocentos metros de altura, computadores mais finos que uma folha de papel que reagem ao pensamento humano e tantas outras coisas dignas da ficção científica. Mas nesse mundo de fantasia futurista, todos vão falar exatamente como falam hoje. E isso é simplesmente (e humanamente) impossível.

A reação à mudança linguística é um traço universal das culturas humanas. A língua está de tal forma entranhada em cada um de nós que imaginar que ela um dia deixará de ser o que é se revela uma ideia insuportável, uma noção capaz de causar em muitas pessoas, mesmo que inconscientemente, um medo quase semelhante ao medo de morrer. Porque a mudança linguística é, de fato, a morte da língua tal como uma geração de falantes a conhece (muito embora a língua esteja também, a todo instante, além de morrendo, renascendo). Não é por outra razão que, em praticamente todas as sociedades, se verifica o lamento, da parte dos mais velhos, contra o "descaso" das gerações mais novas com relação à língua, contra a suposta "pobreza" da fala dos mais jovens. Qualquer investigação rápida de textos escritos com muitas décadas ou mesmo séculos de distância por pessoas que assumem a tarefa de "defender" a língua mostra que o discurso dessa defesa é sempre o mesmo, os argumentos se repetem de uma época para a outra, inalterados, de modo que uma coluna de jornal escrita sobre o tema em 1910 poderia ser estampada em 2010 quase sem alterações. Os linguistas ingleses James e Leslie Milroy batizaram esse traço cultural de "tradição da queixa". E William Labov fala da saudade que as pessoas têm de uma "Idade de Ouro" da língua, em que ninguém cometia "erros", uma época mítica que, por isso mesmo, jamais existiu nem existirá.

A percepção da mudança linguística foi um dos principais impulsos para a criação da disciplina gramatical, no século III a.C., na cidade egípcia de Alexandria, que

era então o mais importante centro de cultura grega. Os filólogos que trabalhavam na famosa Biblioteca de Alexandria perceberam que a língua grega registrada nos documentos antigos, como nos poemas de Homero (século VIII a.C.) era muito diferente da língua grega falada no seu cotidiano. Essa percepção do óbvio, no entanto, veio envolvida num manto de avaliação negativa: a língua tinha mudado, sim, mas mudado para pior. Afinal, eles comparavam o incomparável: a língua escrita literária do passado com a língua falada espontânea do presente. Sem reconhecer as especificidades próprias da produção oral e as diferenças entre os gêneros discursivos, eles julgaram que a língua falada era caótica e sem regras. Para resolver o que lhes parecia um problema, criaram todo um aparato de conceitos, categorias de análise e regras de uso da língua, tal como achavam que era a maneira certa de usá-la. Assim nasceu a gramática normativa [▸401].

4.1.2 Visões científicas da mudança

No início do século XIX — isto é, mil e seiscentos anos depois dos filólogos alexandrinos —, a descoberta do parentesco entre o sânscrito, língua já não mais falada, mas preservada na riquíssima literatura religiosa da Índia, e as principais línguas europeias representou uma importante ruptura com o passado dos estudos linguísticos, essencialmente normativos ou então voltados para questões filosóficas. Como era possível que línguas faladas em lugares tão distantes como a Irlanda e a Índia pudessem ter semelhanças tão grandes?

SÂNSCRITO	IRLANDÊS ANTIGO	SIGNIFICADO
dāru	daur ("carvalho")	"madeira"
dvā(u)	da	"dois"
dvāras	dorus	"porta"
mádhya-	mid	"meio"
mātár-	máthir	"mãe"
náva-	nūe	"novo"
rudhirá	ruad	"vermelho"
sal-ilá	salann	"sal"
sneha-	snechta(e)	"neve"
t(u)vám	tú	"tu"
tráyas	trí	"três"

Esse movimento de comparação entre as línguas para determinar se pertenciam a uma mesma família fez surgir a filologia comparada, que teve um formidável florescimento no século XIX, com especial destaque para os alemães. Graças a hipóteses muito engenhosas e a métodos de reconstituição de estados passados de várias línguas, foi possível estabelecer o parentesco entre um grande número de línguas que vieram a ser agrupadas na chamada família indo-europeia. O nome indo-europeu tentava dar conta precisamente do fato dessas línguas aparentadas se estenderem desde o extremo ocidental da Europa até a Índia.

Foram postuladas, então, várias hipóteses sobre quando e onde viveram os "indo-europeus", uma população que falava uma língua que recebeu o nome de *proto-indo-europeu*. Investigações arqueológicas e teorias linguísticas vêm enfrentando, desde então, a tarefa de localizar o lugar e o tempo mais remotos em que essa língua pré-histórica foi falada. O que parece certo é que ela já existia por volta do ano 6.000 a.C. e que em 3.000 a.C. as primeiras diferenciações já deviam ter ocorrido. Os falantes originais da língua certamente se dispersaram em diferentes direções, talvez porque sua população tivesse crescido muito e fosse necessário buscar alimento e abrigo em outras áreas. Uma corrente migratória se dirigiu para a Europa ocidental. Outra corrente desceu até a Turquia atual, passando pelo Cáucaso. E outra corrente partiu em direção ao Oriente, passando pelo Irã (e se fixando em parte ali, já que o persa é uma língua indo-europeia), pela Índia e chegando até a China onde, na província de Xinjiang, foi descoberta no início do século XX a língua tocariana [▸277].

O estudo comparativo das línguas indo-europeias permitiu conhecer melhor os processos de mudança linguística, principalmente nos níveis fonológico e morfológico. Os cientistas já não se preocupavam com pureza e defesa das línguas, mas com a investigação sobre como e por que as línguas mudam.

A filologia comparada (ou linguística histórica) sem dúvida representa o primeiro grande passo para a constituição de uma ciência das línguas, na concepção moderna de ciência. Um de seus praticantes mais destacados foi o suíço Ferdinand de Saussure (1857-1913). Com base em seus profundos conhecimentos de idiomas antigos e modernos e em suas reflexões sobre o funcionamento das línguas, Saussure propôs uma teoria geral da linguagem, uma linguística geral, com postulados que valessem para todas as línguas humanas, uma linguística que pudesse se impor como ciência autônoma numa época em que as diferentes disciplinas dedicadas ao estudo do ser humano — sociologia, antropologia, psicologia etc. — estavam se constituindo, delimitando seus objetos de estudo e postulando seus próprios conceitos de análise. Era o auge do positivismo, e as ciências humanas e sociais, nessa primeira fase, se deixaram impregnar por essa corrente de pensamento.

A linguística proposta por Saussure recebeu o nome de estruturalista e se dedicava ao estudo da língua em seu aspecto *sincrônico*, isto é, sem recorrer ao passado (ao

diacrônico), sem fazer as comparações entre línguas, tão características do tipo de investigação dominante em boa parte do século XIX. Usando a metáfora do xadrez, recorrente em seus escritos, Saussure dizia que o importante é estudar o jogo tal como ele se apresenta num dado momento, as relações entre as peças tal como colocadas aqui e agora, e não procurar saber que sequência de lances permitiram que a partida assumisse o aspecto que tem aqui nesse instante. Recorrendo a uma metáfora mais contemporânea, poderíamos dizer que para Saussure a língua era como um programa de computador em sua versão mais nova: as pessoas podem muito bem aprender a usar esse programa, conhecer o funcionamento dele, sem precisar saber o que ele tem de novo ou diferente com relação às suas versões passadas.

O estuturalismo exerceu grande impacto e obteve muito sucesso na primeira metade do século XX. Sua dedicação praticamente exclusiva ao sincrônico fez com que os estudos históricos fossem deixados em segundo plano ou mesmo, em alguns lugares, abandonados. Uma das principais características da corrente estruturalista é sua desconsideração do sujeito: no caso da linguística estrutural, a língua era tomada como uma entidade autônoma, independente da vontade e da ação de seus falantes, podendo ser estudada e analisada como um sistema fechado em si mesmo.

Foi preciso esperar até o final da década de 1960 para que a mudança linguística voltasse a ter um papel destacado nos estudos sobre as línguas. Tentando se contrapor a alguns dos postulados da escola de pensamento linguístico dominante na época — o **gerativismo** fundado por Noam Chomsky (nascido em 1928) —, alguns pesquisadores voltaram a se debruçar sobre as transformações sofridas pelas línguas, dessa vez com um enfoque novo, inspirado nas investigações etnográficas, antropológicas, sociológicas. Entre esses investigadores se destaca **William Labov** (nascido em 1927) [▸62], considerado o fundador e principal representante da **sociolinguística variacionista**.

A contribuição mais importante da sociolinguística é decerto sua firme postulação do caráter intrinsecamente **heterogêneo** de toda e qualquer língua humana viva. Em qualquer sociedade, a língua nunca é um sistema único, fechado em si mesmo, mas um conjunto de subsistemas que a sociolinguística chama de **variedades**. As variedades linguísticas representam diversos estágios da mudança linguística, convivendo ao mesmo tempo dentro de uma dada sociedade. Algumas levam adiante processos de mudança que outras estão apenas iniciando, enquanto outras ainda conservam formas mais antigas, enquanto outras já completaram os processos de mudança. A comprovação empírica disso é muito fácil, bastando comparar as diferentes variedades faladas num dado território.

No Brasil, por exemplo, encontramos comunidades em que os falantes pronunciam como ['lũa] o que a maioria dos brasileiros pronunciam como ['lua], ou seja, *lua*. A pronúncia ['lũa], com seu *u* nasal, representa a conservação de uma pronúncia

medieval, atestada até por escrito, com til sobre *u* e tudo, em textos provenientes dos séculos anteriores à chegada dos portugueses no Brasil. Na formação do galego, o *n* intervocálico de palavras como *lana, luna, leona, corona, arena* sofreu síncope [▸154], isto é, "caiu", mas antes de desaparecer transferiu seu traço nasal para a vogal anterior a ele: assim, encontramos em textos medievais as formas *lãa, lũa, leõa, corõa, arẽa*. Mais tarde, os hiatos resultantes da queda do *n* vão ser eliminados, como no caso de *lã* e *areia*, ou vão permanecer, perdendo no entanto a nasalidade: *lua, leoa, coroa*. Como é fácil perceber, os falantes que dizem ['lũa] conservam uma pronúncia que sem dúvida estava presente na fala dos primeiros colonizadores portugueses. Embora a pronúncia ['lũa] esteja hoje restrita a comunidades rurais do interior do Nordeste, nessa mesma região, nas zonas urbanas e na fala de cidadãos mais letrados, é comum ocorrer pronúncias como ['ũa] (*uma*), ['vẽa] (*venha*), ['tẽa] (*tenha*), com hiato, e que conservam a pronúncia anterior ao surgimento da consoante [ɲ], o que pode levar a crer que essa era a pronúncia vigente entre os primeiros portugueses que chegaram por aqui.

Por outro lado, esses mesmos falantes que pronunciam ['lũa] apresentam um traço fonológico que representa um avanço, uma inovação. É a pronúncia [ɪ] para o que escrevemos com *lh*, como *trabalho* que esses falantes pronunciam [tra'baɪu]. Esse "passo adiante" também foi dado pelos falantes de francês, de modo que o grupo escrito *-ill-*, que representava antigamente a consoante [ʎ], hoje se pronuncia como se os dois L não estivessem ali: *famille* [fa'miɪə], *abeille* [a'bɛɪə], *fille* ['fiɪə] ('família', 'abelha', 'filha'). Observe que a pronúncia dessas palavras em francês padrão se assemelha ao modo como muitos brasileiros pronunciam as palavras *família, abelha* e *filha*.

A pronúncia [ʎ] dessas palavras representa a conservação, por parte dos falantes urbanos mais letrados, de um estágio antigo da língua. Em várias línguas do grupo espanhol, a consoante [ʎ], que se escreve como *ll*, passou por diversas transformações: no castelhano padrão da Espanha essa consoante se conserva, mas em outras línguas do grupo ela se realiza como [ɪ] (no andaluz e no peruano, por exemplo), como [ʒ] (no uruguaio e no argentino-portenho), como [ʃ] (no uruguaio e no argentino-portenho de gerações mais jovens). Também no galego moderno o que se escreve *ll* se pronuncia como [ɻ] ou [j], enquanto a pronúncia [ʎ] está confinada às variedades rurais e principalmente à fala das pessoas idosas. Todas essas transformações ocorridas em outras línguas — e que não sofrem nenhum grau de estigmatização nas sociedades onde as pessoas falam assim — mostram que há razões de ordem articulatória para a evolução da consoante [ʎ] rumo a uma semivogal e daí a uma nova consoante palatal. A pesada carga de discriminação que incide no Brasil sobre os falantes que pronunciam [tra'baɪu], ['paɪa], [a'beɪa] etc. é de ordem estritamente social e nada tem a ver com alguma suposta incapacidade dessas pessoas de "falar direito".

Os exemplos acima podem levar a pensar que esses fenômenos se restringem às variedades rurais ou das pessoas sem escolarização formal. Mas isso é um equívoco.

Em todas as variedades do PB, podemos encontrar esses exemplos de conservadorismo e inovação presentes no cotidiano da língua.

Em todas as variedades do PB, incluindo as dos falantes urbanos altamente letrados, as construções com verbos causativos e sensitivos [▸593] se realizam de um modo inovador, que diferencia o PB de todas as outras línguas da família românica. No lugar dos pronomes oblíquos (*me* e *o* principalmente), empregamos os pronomes retos: *deixa* eu *falar, vi* ele *entrar, espera* ela *chegar, o diretor mandou* eu *refazer o relatório* etc. Outra característica exclusiva do PB com relação às outras línguas da família é o uso do verbo na não-pessoa do singular [▸818] para expressar indeterminação do sujeito: *aqui* faz *fotocópia, não sei onde* compra *os ingressos, nessa vaga não* pode *estacionar*, tá usando *muito esse comprimento de vestido* etc.

Por outro lado, o PB, incluindo as variedades urbanas cultas, também apresenta traços conservadores que diferenciam ele, por exemplo, do português europeu, como o uso de *ele* na função de objeto direto [▸797], tal como acabei de escrever, e do verbo *ter* com sentido existencial [▸620]. Esses mesmos usos ocorrem no português angolano e moçambicano, o que indica que também ocorriam no português europeu antigo. O abandono de *ele* como objeto direto e de *ter* com sentido existencial representa uma inovação que se deu no português europeu e só lá. Textos medievais comprovam que nessa língua esses usos também já foram comuns. Como, infelizmente, a norma-padrão até hoje se inspira nos usos dos portugueses — o que é um rematado absurdo sob todos os pontos de vista —, existe o patrulhamento e a repressão injustificada contra tais usos.

4.1.3 Diacronia na sincronia

O que o trabalho dos sociolinguistas mostrou é que a língua não se transforma por inteiro e de uma vez só: como a sociedade é heterogênea, as diferentes comunidades de fala que a compõem apresentam diferenças no uso da língua. Essas diferenças se devem à história de cada comunidade. A sociedade brasileira não está toda ela vivendo segundo tendências, crenças, ideias e práticas iguais para todos os cidadãos. Em regiões mais afastadas dos grandes centros urbanos é possível encontrar pessoas vivendo sem energia elétrica, sem água encanada, sem rede de esgoto, sem televisão, sem escola, sem acesso à internet. É bastante previsível que ali as pessoas falem de um modo que se distancia grandemente das variedades urbanas e que empreguem palavras e expressões antigas que já não são empregadas pelos falantes urbanos, além de também usarem formas novas, desconhecidas das demais comunidades de fala.

E, mesmo nas grandes cidades, as profundas diferenças socioculturais e socioeconômicas que caracterizam a sociedade brasileira podem revelar profundas diferenças

entre os modos de falar dos diferentes grupos sociais. As variedades das periferias da cidade de São Paulo, por exemplo, se distinguem em muitos aspectos das variedades dos bairros tradicionais da cidade.

A organização social de Brasília, com um Plano Piloto dominado pela classe média com a maior renda *per capita* do país e uma constelação de cidades-satélites reservadas para as camadas mais pobres, reproduz em larga escala e com toda a complexidade contemporânea a antiga separação entre casa-grande e senzala dos tempos coloniais. A proporção de habitantes de origem nordestina, por exemplo, é muito maior nas cidades-satélites do que no Plano Piloto, reproduzindo na escala do Distrito Federal a divisão do Brasil entre uma macrorregião mais rica e desenvolvida (Centro-Oeste/Sudeste/Sul) e uma macrorregião mais pobre e menos desenvolvida (Norte/Nordeste). Por sua própria história, Brasília representa um excelente laboratório para os estudos sociolinguísticos, como os que tentam detectar de que modo a procedência regional dos habitantes influencia a formação de uma fala tipicamente brasiliense, depois que se completou, em 2010, o primeiro meio século de vida da capital.

Apesar do nivelamento das diferenças sociolinguísticas que presenciamos na história recente do Brasil [▸248], as marcas mais características das variedades regionais e sociais permanecem, como elementos definidores das identidades individuais e dos grupos, além de servirem como instrumentos para as práticas de discriminação/valorização e de exclusão/privilegiamento tão frequentes em nossa sociedade.

Com a sociolinguística, ficou revelado que todo momento na vida de uma língua (ou melhor, na vida de uma comunidade de falantes, já que a língua não tem vida própria, independente das pessoas que a falam) é atravessado por forças de conservação e de inovação devidas à heterogeneidade social. Em vez de continuar abordando o fenômeno linguístico segundo a dicotomia *sincronia* vs. *diacronia*, instaurada pela linguística estruturalista, a sociolinguística propõe a noção de *diacronia na sincronia* ou pancronia.

Para comprovar o convívio simultâneo dos estágios de mudança linguística numa comunidade, a sociolinguística propõe uma metodologia de pesquisa chamada tempo aparente. Ela consiste no estudo dos modos de falar de diferentes gerações de falantes. Partindo da ideia de que o modo de falar de um indivíduo se consolida no final da adolescência e início da vida adulta, os sociolinguistas distribuem uma dada comunidade em diversas faixas etárias e investigam a realização de determinadas regras (fonológicas, morfossintáticas etc.), de modo a verificar o que mudou de uma geração para outra. Quando se compara os usos ao longo de uma faixa de tempo, o estudo é chamado de tempo real. Dizemos então que é possível detectar *mudanças em tempo aparente* e *mudanças em tempo real*. O exemplo abaixo mostra isso com relação ao uso dos demonstrativos [▸791] na fala da cidade do Rio de Janeiro num intervalo de vinte anos:

tempo real →

tempo aparente ↓

	AMOSTRA 1980							AMOSTRA 2000						
	IDADE	ESTE	ESSE	AQUELE	ISTO	ISSO	AQUILO	IDADE	ESTE	ESSE	AQUELE	ISTO	ISSO	AQUILO
	07-14	01	103	62	00	74	05	07-14	00	188	43	00	89	01
	15-25	00	547	292	00	339	49	15-25	00	343	136	00	443	57
	26-49	02	509	219	00	430	85	26-49	02	303	103	00	322	53
	+50	04	491	254	00	328	102	+50	04	318	102	02	326	53
	TOTAL: 3.896	07	1.650	827	00	1.171	241	TOTAL: 2888	06	1.152	384	02	1.180	164
		0,17%	43,35%	21,22%	00	30,05%	6,18%		0,20%	39,88%	13,29%	0,06%	40,85%	5,6%

[Fonte: Cláudia Roncarati, "Os mostrativos na variedade carioca falada", in: Paiva & Duarte (orgs.) (2003: 143).]

Como os dados deixam ver, o antigo sistema ternário do português clássico — *este/esse/aquele* — virtualmente desapareceu da variedade estudada, tendo sido substituído por um sistema binário — *esse/aquele* — que, como vamos ver em [▸791], compensa a perda do *este/isto* combinando os demonstrativos com advérbios de lugar: *esse aqui, esse aí / isso aqui, isso aí*, combinação que também se faz com *aquele*: *aquele ali ~ lá / aquilo ali ~ lá*. O que vale para a variedade carioca falada também vale para todo o PB falado no Brasil e já tem ampla repercussão nos gêneros escritos monitorados, de modo que é certo dizer que ocorreu uma mudança, já plenamente instalada, no sistema demonstrativo da nossa língua.

O grande achado teórico da sociolinguística foi o conceito de variação ordenada, em oposição ao de *variação livre* das escolas anteriores. Através de investigações empíricas — inserção do pesquisador na comunidade, coleta de dados reais da língua falada por meio da gravação, análise dos dados segundo métodos cada vez mais sofisticados — foi possível demonstrar o papel fundamental da variação nos processos de mudança. Vamos falar dele mais adiante.

4.2 Mudança linguística: um processo sociocognitivo

A mudança linguística é um fenômeno complexo que sempre tem desafiado os estudiosos. A primeira reação foi a que comparava a mudança à "ruína" e à "corrupção" da língua. Essa atitude permanece muito enraizada até hoje na maioria das sociedades. Mas já faz um bom tempo que a pesquisa científica acerca da mudança linguística abandonou essas concepções e tem oferecido respostas mais racionais e interessantes para a pergunta sobre como e por que as línguas mudam.

Com o que já sabemos até o momento, é possível dizer que a mudança linguística é um processo social e cognitivo. Isso significa que dela participam fatores socio-

culturais, decorrentes das dinâmicas de interação dos indivíduos e das populações de uma dada comunidade, e fatores sociocognitivos, derivados do funcionamento do nosso cérebro quando processamos a língua que falamos (e fazemos isso a cada segundo), processamento que implica não só o indivíduo, como também os demais com quem ele interage.

O cérebro não está fechado em si mesmo, não é a "caixa preta" dos behavioristas do início do século XX: ele se encontra num corpo que, por sua vez, está mergulhado inexoravelmente na vida social, em constante interação com outros. A língua também não é o "mentalês" postulado pelo gerativismo chomskiano, que despreza tudo o que não for "de dentro para dentro". Ou seja, a mudança linguística, como tudo o que se refere ao processamento mental, decorre de uma elaboração conjunta, de uma atividade coletiva. Nenhuma pessoa pode mudar sozinha o que quer que seja na língua. As mudanças só se implementam porque encontram repercussão cognitiva no cérebro dos falantes da língua e aceitação na comunidade. Por isso, adotamos aqui a tese de que a mente humana age segundo suas estruturas biológicas, sim, mas que sua formação, formatação e configuração é um processo essencialmente social. Da interação dessas duas ordens de fatores, socioculturais e sociocognitivos, resulta a mudança linguística.

Os fatores socioculturais são estreitamente dependentes da vida de cada comunidade de fala: o tamanho de sua população, a distribuição dos papéis sociais atribuídos às mulheres e aos homens, sua interação mais ou menos intensa com outras comunidades de fala, sua matriz econômica e produtiva, sua hierarquia social, seu sistema político, suas instituições, seu sistema de ensino, seu grau de desenvolvimento tecnológico, sua maior ou menor integração às redes de comunicação, de intercâmbios culturais, tecnológicos e comerciais nos níveis local, nacional e global, a composição étnica e etária dessa comunidade, suas relações ecológicas com o território que habita etc. Como todos esses fatores são altamente variáveis de um lugar para o outro, os processos de mudança linguística decorrentes deles também serão altamente variáveis de um lugar para o outro, mesmo no interior de uma sociedade considerada globalmente uma só, como a brasileira.

Os fatores sociocognitivos que estimulam a mudança linguística já são de outra natureza. Dado que todos os seres humanos compõem uma única espécie e dispõem dos mesmíssimos recursos intelectuais, da mesmíssima potencialidade cognitiva, do mesmíssimo cérebro e da mesma configuração fisiológica, é mais do que esperável que, no processamento da língua, surjam fenômenos de mudança comuns a todos os idiomas do mundo. E isso de fato ocorre, como pesquisas com números cada vez maiores de línguas têm demonstrado fartamente, de modo que é possível falar da existência de tendências universais de mudança linguística. Esses universais, no entanto, não são de natureza exclusivamente genética, inata, como postula, por exemplo o gerativismo de Chomsky. A interação dessas tendências sociocognitivas

universais com os fatores ecossocioculturais presentes na comunidade é que vão determinar os ritmos da mudança, que diferem de uma sociedade para outra e, dentro de uma mesma sociedade, de uma comunidade para outra.

Esses fatores de ordem sociocognitiva constituem forças centrífugas no processo de mudança, isto é, forças que agem de modo que a língua se afaste cada vez mais do que é para se tornar o que será.

Já os fatores sociais que intervêm na mudança linguística, eles podem constituir forças centrífugas ou forças centrípetas. Os fatores sociais *centrífugos* são primordialmente a variação e o contato, e vamos tratar deles logo adiante. As forças sociais centrípetas são aquelas que procuram manter a língua na maior estabilidade e imutabilidade possível. É o que a gente examina agora.

4.3 Fatores sociais e mudança linguística

4.3.1 Forças centrípetas: as instituições sociais

No que diz respeito ao ambiente social, é inevitável que existam, na sociedade, forças centrípetas que agem sobre a língua, forças que puxam a língua para o centro, que refreiam ela, que tentam conter seu impulso de mudança. Essas forças são exercidas pelas instituições sociais que, de maneira explícita ou não, oficial ou não, tentam impor algum controle sobre os destinos do idioma.

A mais importante de todas, evidentemente, é a escola, o sistema formal de ensino que, em todos os seus níveis, tenta dar aos cidadãos (ou a alguns deles, no caso de sociedades marcadamente desiguais como a brasileira) uma educação sistematizada, programada de acordo com currículos definidos pelas instâncias oficiais. A escola tenta veicular uma cultura que está geralmente associada com as camadas sociais privilegiadas e, por conseguinte, transmitida na roupagem de uma "língua" considerada "culta" ou "exemplar".

Junto com a escola, outras instituições e agentes sociais também contribuem para o esforço de refrear a mudança linguística:

- o peso, maior ou menor, da tradição literária (que elege alguns escritores como os "clássicos" do idioma, modelos a serem "imitados");
- o trabalho dos gramáticos e dicionaristas, empenhados em descrever e prescrever a língua "certa";
- a burocracia em geral, o sistema jurídico, o poder legislativo (com suas fórmulas estereotipadas e seu fraseado típico, em geral muito rebuscado e repleto de expressões obsoletas);

- todo o aparato estatal com sua multiplicidade de órgãos e serviços públicos;
- as instituições religiosas que, em geral, se apoiam em textos antigos, reverenciados, e que precisam ser mantidos no estado de máxima "pureza" original, para que não sejam "distorcidas" as verdades que foram reveladas aos fiéis pelas forças divinas (o índice de pessoa *vós*, por exemplo, só sobrevive, no PB de hoje, em alguns textos religiosos, como as traduções da Bíblia);
- as academias de língua, nos países em que elas existem e têm uma presença atuante na vida cultural da sociedade (como é o caso, por exemplo, da Real Academia de la Lengua Española, e como nem de longe é o caso da Academia Brasileira de Letras);
- e, mais recentemente na história da humanidade, os poderosíssimos meios de comunicação, que necessitam de uma linguagem mais ou menos uniformizada para exercer suas funções de formação (e sobretudo de *conformação*) da opinião pública.

Um elemento fundamental em tudo isso é, sem dúvida alguma, a presença da escrita. Tudo o que tem a ver com a instituição de uma linguagem "certa", "oficial", "uniforme", "normatizada" etc. também tem a ver com o uso intenso da *escrita*. As sociedades que são fortemente grafocêntricas, isto é, em que a *cultura escrita* é onipresente e supervalorizada, são também aquelas que ostentam instituições com grande poder *centrípeto* sobre as forças de mudança da língua. Essa concepção da escrita como algo superior vem desde a Antiguidade clássica, e a valorização da língua falada como verdadeiro objeto de estudo das ciências da linguagem só veio a ocorrer em meados do século XIX e, de forma mais explícita e sistematizada, no início do século XX, com o estruturalismo. Evidentemente, sabemos que em sociedades desprovidas de escrita, de cultura letrada e de escola formal, também há forças centrípetas de regulação linguística, como a distribuição do poder, as hierarquias ocupadas pelos diferentes membros da comunidade, os papéis reservados a homens e mulheres, o prestígio dos idosos, reconhecidos como detentores de maior sabedoria etc. No entanto, o que nos interessa aqui é o papel das instituições em culturas mais amplas e complexas, em que existe uma cultura de prestígio fortemente influenciada pela escrita e pela escolarização formal.

Assim, é fácil concluir que quanto menor for a presença e a influência da escrita institucionalizada (o que significa menor presença/influência da escola, do poder do Estado, dos meios de comunicação etc.), maior e mais rápida também será a atuação das forças *centrífugas*, socioculturais e sociocognitivas, que favorecem a mudança das línguas. A pressão da escrita sobre a língua falada pode ser responsável pela criação de um imaginário linguístico popular, onde a língua falada tem de ser um reflexo da língua escrita (o que, naturalmente, é impossível, dadas as limitações do sistema de notação de qualquer ortografia para dar conta de todos os fenômenos de variação presentes na língua falada). Também pode provocar mudanças que, de outra maneira, não ocorreriam.

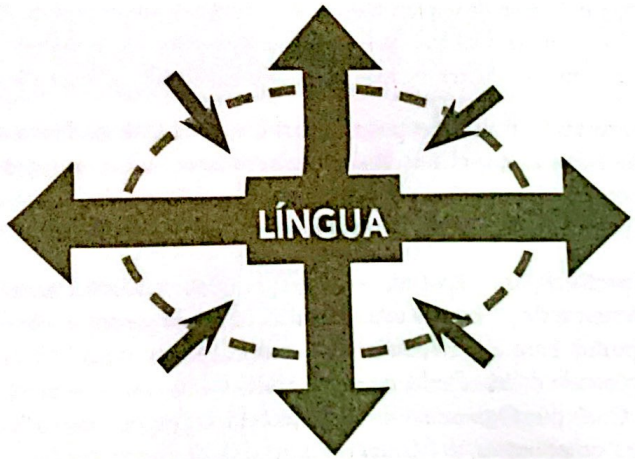
Por exemplo, a pronúncia da consoante *m* na palavra *uma* é decorrência exclusiva de sua forma escrita: sendo o masculino *um* [ũ], o feminino natural é ['ũa], como é

a pronúncia corrente entre falantes de algumas variedades regionais (e também em galego), mas por caprichos da ortografia, o feminino se formou com o acréscimo de um *-a* à forma do masculino, em que o *m* é só um índice de nasalidade da vogal e não uma consoante. Disso resultou a pronúncia ['ũma], calcada na ortografia. Como se pode ver, a fixação da escrita, mesmo agindo como força centrípeta contra as mudanças, acaba provocando mudanças imprevistas.

O conservadorismo da ortografia, ou seja, da escrita como instituição gerida pelo Estado, chega a extremos em certas línguas, como o francês. A forma escrita oficial dessa língua retrata em boa medida a pronúncia do francês no século XVII. Apesar disso, e da pressão normativista característica da cultura linguística francesa, as mudanças ocorridas na língua falada provocaram um distanciamento radical entre o que se escreve e o que se fala, tornando o aprendizado da escrita francesa uma tarefa exigente e demorada:

- BEAUCOUP [bo'ku] ('muito')
- EAUX [o] ('águas')
- ILS PARLAIENT [il par'lɛ] ('eles falavam')
- OISEAUX [ʊa'zo] ('pássaros')
- ROULEAUX [ru'lo] ('rolos')

Essas forças centrípetas, portanto, que partem das instituições que tentam cercear a língua (representadas no desenho abaixo pelo perímetro tracejado) conseguem somente conter ou atrasar por algum tempo a mudança linguística (por isso, no desenho, elas estão representadas como setas bem menores do que as forças centrífugas inerentes aos processos cognitivos). Elas jamais terão o poder de impedir totalmente nem (muito menos) para sempre essa mudança porque, sendo de natureza sociocognitiva, a mudança é muito mais poderosa do que qualquer outra força social institucionalizada.



4.3.2 *Forças sociais centrífugas: a variação*

A sociolinguística variacionista surgiu do esforço de alguns pesquisadores que tentavam encontrar novas respostas para o processo de mudança. Para a linguística estruturalista clássica, o importante era conhecer as *estruturas* de funcionamento da língua, definir principalmente os elementos que se encontravam em *oposição*, de modo a ser possível construir a abstração chamada *sistema* [▶51]. Seu principal foco de interesse foram a fonética-fonologia e a morfologia.

Para uma abordagem estruturalista clássica, é de menor importância o reconhecimento de que existem brasileiros que pronunciam a palavra *porto* como ['portu], ['poɹtu], ['poɾtu], ['pohtu], ['poxtu] etc. O mais importante é reconhecer que todas essas diferentes pronúncias da consoante [r] são realizações empíricas, audíveis, de uma entidade teórica, chamada fonema, que se opõe a outras entidades teóricas no nível da fonologia da língua. Para delimitar o fonema, os estruturalistas recorriam à semântica: se todas essas pronúncias (se todos esses *significantes*) remetem a um mesmo *significado*, 'lugar onde as embarcações atracam', então elas são *alofones* (sons diferentes) de um fonema (que, convém insistir, não é um som, é uma entidade abstrata, teórica). Se o [r] fosse trocado, digamos, por um [s], teríamos ['postu], ou seja, *posto*, com outro significado. A conclusão é que [r] e [s] são fonemas do PB, porque permitem a diferenciação das palavras. Sendo assim, não podem ser intercambiados, não podem aparecer um no lugar do outro: eles se encontram em *distribuição complementar*. Desse modo, é possível dizer que no PB existe um fonema /r/ que se realiza como [r, ɹ, ɾ, h, x].

Essa postura teórica e metodológica se revelou muito útil para o estudo de línguas indígenas sem tradição escrita e ainda não descritas. Nos Estados Unidos, na primeira metade do século XX, os linguistas estavam muito preocupados em conhecer as muitas línguas indígenas de seu território, antes que elas desaparecessem. Para saber quantos e quais eram os fonemas de uma língua desconhecida, o método de comparar palavras com sons diferentes, mas de mesmo significado, parecia adequado.

O fato de haver tantas realizações possíveis para o fonema /r/ de *porto* era atribuído pelos estruturalistas ao que chamavam de "variação livre". A sociolinguística vai se opor a esse conceito, postulando que a variação é condicionada, seja por fatores sociais, seja por fatores estruturais, linguísticos.

No caso de ['portu], ['poɹtu], ['poɾtu], ['pohtu], ['poxtu] etc., o condicionamento é *social*: essas diferenças de pronúncia estão vinculadas à *origem geográfica* do falante. A pronúncia ['poɹtu], com o "R caipira", é característica, por exemplo, dos moradores do interior do estado de São Paulo, do sul de Minas Gerais, do norte do Paraná, do centro-sul de Goiás etc. O mesmo vale para a pronúncia ['poʃtu], com o "s chiado", que caracteriza, entre outros, os falantes cariocas.

Outro fator social muito importante é o grau de escolarização dos indivíduos nos usos de determinadas formas linguísticas. A pressão/repressão normativa da escola acaba exercendo grande influência no emprego e manutenção das formas linguísticas socialmente mais prestigiadas, como se vê abaixo, num estudo sobre a concordância nominal na fala carioca, em que, quanto mais alto o nível de escolarização, maior a frequência de realização das marcas de concordância nominal:

ESCOLARIZAÇÃO	FREQUÊNCIA	%
Primário	950/2.398	40
Ginásio	1.297/2.664	57
2º grau	1.499/2.082	72

[Fonte: M. M. P. Scherre, "Sobre a influência de variáveis sociais na concordância nominal", in: Silva & Scherre (orgs.) (1996: 242).]

As muitas pesquisas sociolinguísticas realizadas no Brasil nos últimos quarenta anos mostram que o fator social escolarização é o mais importante na diferenciação das variedades linguísticas em nosso país. O poder nivelador da escolarização provoca em muitas pessoas a substituição de usos linguísticos mais estigmatizados por outros que não sofrem discriminação por parte das camadas mais letradas da população. Com isso, as pessoas que receberam educação formal e têm origem em camadas sociais desfavorecidas apresentam modos de falar diferentes dos de seus pais e demais familiares analfabetos ou semianalfabetos. Embora exista, da parte de muitos linguistas e educadores, a persistente declaração de que a escola não deve "substituir" um modo de falar pelo outro, essa substituição decorre naturalmente da vontade do indivíduo que, ao se ver discriminado (junto com sua comunidade) por seu modo de falar, procura fugir dessa ameaça sociocultural se apropriando das variantes linguísticas de prestígio.

O grau de escolarização também é apontado pelas pesquisas sociológicas como o principal elemento impulsionador da ascensão social: os melhores empregos se destinam a quem tem mais estudo. Como o acesso a uma educação de qualidade ainda é extremamente restrito no Brasil, a aquisição das formas linguísticas de prestígio, a elevação do nível de letramento e a ocupação profissional mais bem remunerada também ficam muito restritas, apesar dos grandes e inéditos avanços sociais obtidos na primeira década do século XXI.

Enquanto em outras sociedades a variação é determinada principalmente por outros fatores sociais, como a etnia, a posição hierárquica de mulheres e homens, a origem geográfica do falante etc., na sociedade brasileira é a desigualdade de acesso aos direitos elementares e à cidadania plena, a começar pela educação, a principal barreira que separa as variedades linguísticas. É de se esperar que, com o aumento progressivo desse acesso, as diferenças entre as variedades linguísticas deixem de

ser causadas pela desigualdade educacional e correspondam bem mais a diferenças regionais.

Os fatores estruturais que condicionam a variação são de diversa natureza. Quando o estudo recai sobre elementos sonoros, por exemplo, é o *ambiente fonético* da variável pesquisada que vai condicionar a sua realização. Por exemplo, o ditongo escrito *ei* apresenta duas realizações bem conhecidas no PB: a monotongação [e] e a ditongação [ɐɪ]:

ameixa	[a'meʃa]
beijo	['beʒu]
cabeleireiro	[kabile'reru]
cantei	[kã'tɐɪ]
cheirei	[ʃe'rɐɪ]
beirada	[be'rada]
falei	[fa'lɐɪ]
jeito	['ʒɐɪtu]
meigo	['mɐɪgu]
peidei	[pɐɪ'dɐɪ]
peito	['pɐɪtu]
queijo	['keʒu]
queixo	['keʃu]
sujeito	[su'ʒɐɪtu]

A conclusão é que a monotongação é favorecida pela presença de uma consoante palatal (ʃ e ʒ), como em *ameixa, beijo, queijo, queixo*, ou de uma vibrante simples (r), como em *beirada, cabeleireiro* etc. Diante de outras consoantes ou em final absoluto de palavra, o ditongo [ɐɪ] conserva sua semivogal.

Podemos exemplificar um fator condicionante de variação no plano da *morfossintaxe* com o uso do pronome *ele* como objeto direto. As pesquisas mostram que esse uso tende a ocorrer com maior frequência quando o objeto direto tem o traço semântico [+animado], como em: *Eu pensei que não conhecia* a Eugênia, *mas me lembrei que você tinha me apresentado* ela *no aniversário da Denise*. Observe:

TRAÇO SEMÂNTICO	PRONOME OBLÍQUO		PRONOME RETO	
	QUANTIDADE	%	QUANTIDADE	%
[+animado]	76	78,4	281	92,4
[-animado]	21	21,6	23	7,6
TOTAL	97	100,0	304	100,0

[Fonte: Mª E. L. Duarte, "Clítico acusativo, pronome lexical e categoria vazia no português do Brasil", in Tarallo (org.) (1989: 24).]

O impacto da variação social sobre a mudança linguística foi o principal alvo das investigações empreendidas por W. Labov na década de 1960 e que se tornaram estudos clássicos da metodologia sociolinguística. Sua intenção era provar que para o estudo da mudança de uma língua era imprescindível incluir na pesquisa a dinâmica social da comunidade em foco. Também era seu objetivo mostrar que é possível acompanhar uma mudança linguística em andamento, enquanto ela ocorre, ao contrário do que diziam os linguistas anteriores a ele, que argumentavam que a mudança operava tão lentamente e imperceptivelmente que era impossível acompanhá-la, a não ser comparando dois estágios da língua muito afastados no tempo.

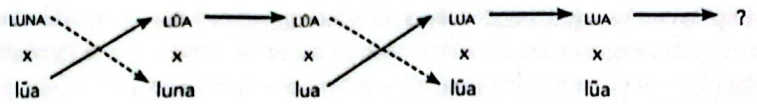
A metodologia do estudo da *mudança em tempo aparente* mostrou que isso não era verdade. Se analisarmos, por exemplo, duas gerações de falantes da variedade portenha (de Buenos Aires) do espanhol argentino, uma mulher de 60 anos e sua neta de 20, podemos encontrar duas realizações diferentes para o dígrafo que se escreve *ll* como nas palavras *llorar, calle, caballo*. A avó pronunciará essa consoante como um [ʒ], de modo que para os ouvidos brasileiros as palavras vão soar como "jorar", "caje", "cabajo". A neta, por sua vez, vai pronunciar a consoante como um [ʃ], de modo que vamos ouvir algo como "xorar", "caxe", "cabaxo". No espaço de uma geração ocorreu uma mudança — o desvozeamento (ou ensurdecimento) da consoante palatal — e os falantes se apercebem disso porque é muito comum que as pessoas mais velhas façam comentários a respeito da fala "errada", "estranha" ou "feia" das mais novas. Com o passar do tempo, a forma nova vai se impor e vai perdurar até que venha a sofrer a eventual concorrência de uma forma ainda mais nova. É assim que as línguas mudam.

Esse tipo de mudança intergeracional ocorre em todos os níveis da língua: na pronúncia, nas regras morfossintáticas, no significado das palavras, na criação de novas palavras etc. Assim, ao contrário de seus predecessores, que acreditavam que a mudança linguística ocorria numa linha reta em que cada segmento da reta era uma transformação imperceptível que se processava, Labov demonstrou que é a competição entre duas ou mais formas em variação numa mesma época e num mesmo lugar que leva à vitória de uma delas, completando uma mudança que, mais adiante, pode sofrer nova concorrência e ser subjugada por uma nova transformação.

Portanto, retomando o exemplo da palavra *lua*, a descrição que geralmente se encontra nas gramáticas históricas assume a seguinte forma:

luna > lũa > lua

Essa representação, no entanto, cria a ilusão de que a mudança ocorre em etapas sucessivas, em que uma forma nova vai substituindo a anterior pela mera passagem do tempo. Além disso, desconsidera o fato de que ainda existem falantes que dizem ['lũa]. Tal como entendida pela sociolinguística, uma mudança ocorre de outra maneira:



Em cada fase da mudança, a forma inovadora se torna majoritária no uso dos falantes daquela comunidade, enquanto a forma antiga sobrevive em grupos sociais mais isolados e/ou na fala dos mais idosos. Com o avanço do tempo, a forma mais antiga desaparece e a forma inovadora se impõe completamente, mas só para sofrer concorrência de outra forma inovadora, e assim por diante. Hoje, no PB, por exemplo existem as duas formas — ['lua] e ['lũa] —, sendo ['lũa] extremamente minoritária e restrita a comunidades isoladas. A difusão da fala urbana pelos meios de comunicação e a expansão do acesso à educação formal podem ser fatores que venham a impor de vez a forma ['lua], fazendo desaparecer a outra, tal como séculos atrás desapareceu ['luna]. Hoje em dia é possível ouvir duas realizações para o que se escreve *lua*: o hiato ['lua] e a pronúncia com uma semivogal ['luʊa], continuação de uma tendência muito antiga à eliminação dos hiatos. Isso já é um indício de uma provável futura mudança.

Além de ocorrer entre as gerações de falantes, a mudança se processa também através da **variação entre as classes sociais**. Os estudos sociolinguísticos mostram que, em geral, uma forma inovadora emerge na fala dos indivíduos das classes média baixa e baixa (D, E) e vai subindo na escala social até ser incorporada pelos falantes das camadas médias altas e altas (C, B, A). Quando ela se instala nesse último grupo social, o prestígio dos falantes se transfere para a inovação linguística e a mudança se completa.

É muito comum que a forma inovadora, quando ainda está restrita aos falantes de menor prestígio social, sofra uma **avaliação negativa** por parte dos grupos socioeconômicos dominantes. Quando essa *avaliação* deixa de ser negativa, é porque parou de ser condenada como "erro" e se tornou plenamente aceita.

No português europeu, por exemplo, mais precisamente na variedade de Lisboa, a vogal /e/ quando precedida de uma consoante palatal (como as que escrevemos com as letras e dígrafos *j, ch, lh, nh* etc.) é pronunciada como [ɐ] (*a* central) resultando em pronúncias como ['fɐʃu], ['vɐʒu], [ʃ'pɐʎu], [i'gɾɐʒɐ], ['sɐʒɐ] para o que se escreve *fecho, vejo, espelho, igreja* e *seja*. Essa mudança fonética (totalmente estranha ao PB) começou a ser notada na primeira metade do século XIX. Um intelectual português, José Inácio Roquete, escreveu em 1845 que essa pronúncia era *"muito comum entre a gente ordinária de Lisboa"*. Já em 1883, outro estudioso português, Gonçalves Viana, escrevia que todos os lisboetas falavam assim e que somente *"algum caturra velho"* conservava a pronúncia antiga. A mudança de avaliação num espaço de quarenta anos revela a plena incorporação da nova pronúncia em todas as camadas sociais, inclusive na mais letrada.

As inovações linguísticas que têm mais chance de se implementar são as que surgem nas camadas médias baixas da população, uma vez que as inovações surgidas nos grupos sociais mais desfavorecidos sofrem avaliação muito mais negativa da parte dos demais grupos.

4.3.3 *Forças sociais centrífugas: o contato linguístico*

Junto com a variação, o contato linguístico é outro fator social importantíssimo para a implementação das mudanças linguísticas. Os fatores sociocognitivos que permitem a dedução de *universais* nas mudanças das línguas atuam em toda e qualquer comunidade, já que se vinculam ao processamento da linguagem no cérebro dos falantes e à construção conjunta da língua. Assim, mesmo a língua falada por uma população muito isolada, sem contato com nenhuma outra sociedade — como tem sido o caso de muitas línguas indígenas brasileiras e de línguas faladas em outros lugares, como na Nova Guiné — vai passar inevitavelmente por processos de mudança. O que ocorre em situações assim é que o ritmo da mudança é mais lento, mais gradual do que em sociedades mais complexas, sujeitas a interações frequentes e de todo tipo com outras sociedades. Por isso, a noção romântica de um "dialeto puro", "intocado", falado em algum lugar remoto do país (como o mito cultural que existe nos Estados Unidos de que os habitantes das montanhas Apalaches "falam exatamente como Shakespeare escrevia", ou no Brasil de que o português mais "puro" é falado no Maranhão) não tem nenhum fundamento científico. Mesmo as comunidades brasileiras mais isoladas, embora apresentem traços conservadores em sua variedade linguística, apresentam traços inovadores resultantes das inevitáveis mudanças.

Quando uma população falante de uma língua A se vê constrangida a falar uma língua B diferente da sua — por emigração, conquista, banimento, colonização, escravização etc. —, o contato dessas duas línguas provocará mudanças principalmente na língua B, devidas a diversos fatores que se conjugam nesse processo, entre os quais a exposição mais ou menos intensa dos novos falantes à língua B, de modo que possam apreendê-la com maior ou menor grau de proficiência. O tipo de constrangimento que leva uma população a tentar se apoderar de uma língua diferente também incide nos processos de mudança. O mais importante a ressaltar é que **o contato linguístico acelera o ritmo das mudanças** que, em outra situação, se processariam decerto mais lentamente. Dependendo da natureza desse contato, essas mudanças podem ser mais radicais ou menos radicais. Vamos tratar aqui dos casos de *conquista* e da *colonização/escravização*, importantes para a história da nossa língua.

4.3.3.1 Causas de contato: a conquista

Quando um território habitado por um ou mais povos é invadido e conquistado por outro povo, é possível observar três tipos de situação:

I. o povo conquistador impõe sua língua ao povo conquistado, que a adota e em pouco tempo abandona sua língua ancestral;
II. o povo conquistador adota a língua do conquistado e abandona sua língua ancestral;
III. o povo conquistador nem impõe sua língua nem adota a do povo conquistado.

As línguas que se encontram em jogo nessas três situações distintas de seu *status* sociocultural e político recebem nos estudos histórico-linguísticos as seguintes classificações:

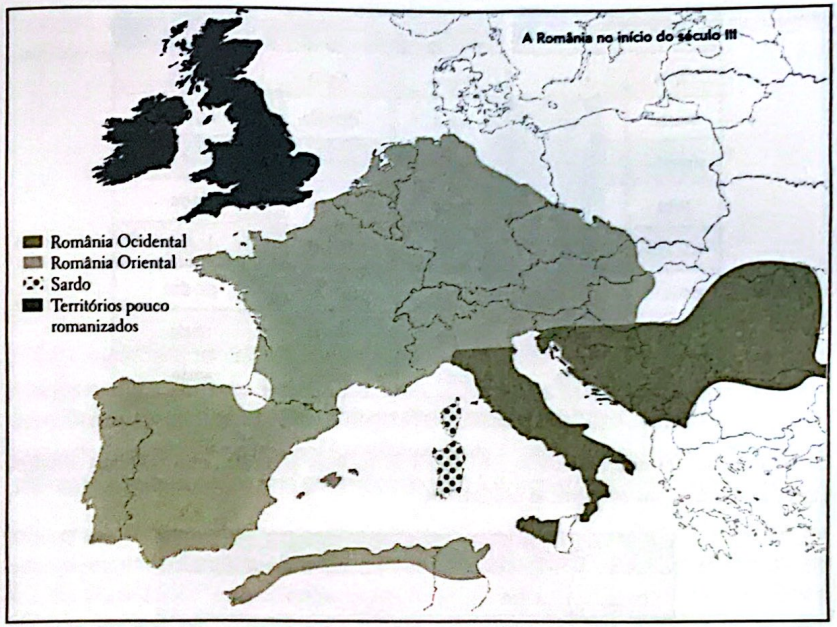
- estrato → é a língua do povo conquistador que é adotada pelo povo conquistado;
- substrato → é a língua do povo conquistado que desaparece, mas influencia o estrato nos níveis fonológico e morfossintático, mas com pouca contribuição lexical;
- superestrato → é a língua do povo conquistador que não é imposta aos conquistados, mas deixa contribuições no estrato, basicamente no nível lexical;
- adstrato → é a língua do povo conquistador que não adota a língua dos conquistados nem impõe a sua; aqui também a contribuição ao estrato é basicamente lexical.

Essas quatro situações ocorreram na Península Ibérica ao longo de mais de mil e quinhentos anos de história. É o que vamos ver no próximo capítulo. Por ora, vamos nos deter na noção de substrato, porque ela é importante para a compreensão da mudança linguística originada por contato de línguas.

4.3.3.2 O substrato céltico

O povo que habitava o extremo noroeste da Península Ibérica quando os romanos conquistaram a região falava uma língua da família céltica. Vamos chamar essa língua de galaico, porque aquela região foi designada pelos romanos como *Gallaecia*, nome que está na origem de Galiza (ou Galícia) e do povo *galego*.

A expansão romana pela Europa acabou impondo o latim como língua de muitos povos diferentes e de um amplo território. Os romanistas, isto é, os que se dedicam ao estudo das línguas românicas — as línguas derivadas do latim — costumam dividir a área linguística ocupada por essas línguas em duas porções: a Românía Ocidental e a Românía Oriental. A Românía Ocidental inclui a Península Ibérica, a França, a Suíça e o norte da Itália (acima da linha imaginária que liga as cidades de La Spezia e Rimini). A Românía Oriental inclui o resto da Itália e a Romênia — antigamente incluía também a língua românica falada na Dalmácia (região que hoje pertence à Croácia, na antiga Iugoslávia), cujo último falante morreu no ano de 1898.



Uma diferença marcante entre as línguas românicas *ocidentais* e as línguas românicas *orientais* está no que aconteceu com as consoantes quando ocorriam *entre vogais*.

Quando em latim as consoantes intervocálicas eram duplas (e pronunciadas como duplas), nas línguas românicas ocidentais elas se tornaram simples:

LATIM	ITALIANO	CATALÃO	ESPANHOL	PORTUGUÊS
admittere	ammettere	admetre	admitir	admitir
battuere	battere	batre	batir	bater
bucca	bocca	boca	boca	boca
opponere	opporre	oposar	oponer	opor
supportare	sopportare	soportar	soportar	suportar
uacca	vacca	vaca	vaca	vaca

[Obs.: O italiano foi incluído para efeito contrastivo: nessa língua, as consoantes duplas são pronunciadas, como eram em latim.]

Já as consoantes *surdas* (produzidas sem participação das *pregas vocais*, como *p*, *t*, *k*) se transformaram nas correspondentes consoantes *sonoras* (com participação das *pregas vocais*, como *b*, *d*, *g*):

LATIM	ITALIANO	CATALÃO	ESPANHOL	PORTUGUÊS
aqua	acqua	aigua	agua	água
acutu	acuto	agut	agudo	agudo
apertu	aperto	obert	abierto	aberto
fata	fata	fada	hada	fada
iocare	giocare	jugar	jugar	jogar
potere	potere	poder	poder	poder
rota	rota	roda	rueda	roda
vita	vita	vida	vida	vida

[Obs.: O italiano foi incluído para efeito contrastivo.]

Se as consoantes intervocálicas em latim já fossem sonoras, elas sofriam *síncope* (queda) nas línguas românicas ocidentais:

LATIM	CATALÃO	ESPANHOL	PORTUGUÊS
pede-	pe	pie	pé
fede-	fe	fe	fé
rege-	rei	rey	rei
lege-	llei	ley	lei
magis	més	más	mais
sedere	ser	ser	ser

O francês foi ainda mais longe nesse processo. Nessa língua as consoantes sonoras surgidas de uma surda latina também sofreram síncope:

LATIM	PORTUGUÊS	FRANCÊS
rota	roda	roue [ru]
vita	vida	vie [vi]
totu-	todo	tout [tu]
lupu-	lobo	loup [lu]
fata	fada	fée [fe]
datu-	dado	dé [de]
amicu-	amigo	ami ['ami]

De modo muito interessante, em catalão, em castelhano, em galego e em português europeu, as consoantes sonoras [b], [d] e [g], quando estão entre vogais (ou antes de

[r] e [l]), são articuladas como fricativas: [β], [δ] e [γ], um fenômeno articulatório que não ocorre no PB:

PORT. BRASILEIRO	CATALÃO, CASTELHANO, GALEGO, PORT. EUROPEU
sa[b]er	sa[β]er
po[d]er	po[δ]er
ce[g]o	ce[γ]o
ca[b]er	ca[β]er
pa[g]ar	pa[γ]ar
vi[d]a	vi[δ]a

É fácil supor que as consoantes sonoras intervocálicas do francês antigo, antes de desaparecerem, tenham passado por uma fase em que tiveram pronúncia africada semelhante ao que ocorre atualmente em catalão, castelhano, galego e português europeu. É o que se verifica em castelhano, onde é muito comum que a terminação *-ado* (particípio passado) seja pronunciada sem o *d*: *canta'o, busca'o, habla'o* etc.

Esse processo em que as consoantes intervocálicas vão sendo substituídas por outras com pronúncia mais "suave" recebe o nome técnico de lenização (ou abrandamento), do latim *lenis*, 'macio, doce, suave, brando'. Ora, os linguistas que estudam a história das línguas célticas nos ensinam que, nessas línguas, a lenização ocorreu sistematicamente ao longo do tempo, num percurso que podemos representar assim, usando *t* como exemplo e palavras do castelhano:

-tt-	→	-t-	→	-d-	→	-ð-	→	-Ø-
0		1		2		3		4
	gutta > gota		vita > vida		dado > daðo		daðo > da'o	

Na formação das línguas românicas ocidentais, esse processo afetou intensamente o latim vulgar falado naquelas regiões. Na antiga Gália (a França atual), a lenização foi mais radical, chegando até o estágio 4 (zero). No catalão, no castelhano, no galego e no português europeu, o processo ainda está na sua fase 3, pré-zero. Se vai chegar à etapa do zero total, só o tempo vai dizer.

O exame dos processos de lenização na história das línguas célticas e sua presença marcante na passagem do latim para as línguas românicas ocidentais levou muitos estudiosos a afirmar que essa diferença no tratamento das consoantes intervocálicas na România Oriental e na România Ocidental se deve precisamente ao fato das regiões ocidentais do Império Romano terem sido habitadas por diversos povos de etnia céltica. No caso da Gália (atual França), a população celta era numerosa e dominava praticamente todo o território. Na Hispânia (Península Ibérica), além dos celtas, que eram maioria, tinha outros povos: iberos, fenícios, gregos, tartessos, que falavam línguas muito diferentes entre si e muito diferentes das línguas célticas. Ao adotar o latim, os povos de língua celta teriam transferido para a nova língua hábitos articulatórios e características morfossintáticas de sua língua ancestral.

Podemos entender um pouco desse processo prestando atenção ao que ocorre quando uma pessoa ou um grupo de pessoas se muda para um país diferente do seu e tem de aprender e usar a língua do país anfitrião. Se um falante argentino, por exemplo, se muda para o Brasil — e se não for uma pessoa especialmente dotada para falar línguas estrangeiras, que é o caso da maioria —, podemos esperar que ele pronuncie a palavra *casa* como ['kasa] e não como ['kaza], uma vez que o som [z] não existe em sua língua, o que vai levá-lo a dizer [a'saguas] para *as águas*, em vez de [a'zaguas] como nós falamos; que não faça uma distinção precisa entre [b] e [v], porque nas línguas do grupo espanhol (e também no galego) essas duas consoantes se confundem, com predominância de [b] mesmo nas palavras escritas com *v*, como *vivo*, que para nós soa parecido com ['bibo]; que insira em seu modo de falar o PB características morfossintáticas próprias de sua língua, como o emprego de um pronome objeto indireto antes da enunciação desse objeto ("*Já lhe* disse *a Pedro* que me esperasse") ou a inversão da ordem sujeito-verbo em orações relativas ("Não vi ainda o carro que *comprou Pedro*").

Algumas características regionais do PB se devem ao contato de línguas europeias que foram trazidas para cá por imigrantes alemães e italianos. Em zonas da região Sul de colonização alemã, é comum, nas variedades de PB ali faladas, não haver distinção entre *carro* e *caro*, pois em alemão não existe diferença entre [r] e [ɾ̌]. Esses mesmos falantes tendem a pronunciar como [ʃ] a consoante [ʒ], de modo que *queijo* e *queixo* têm pronúncia igual: a consoante [ʒ] não existe em alemão. Como esses falantes empregam na vida familiar diferentes dialetos alemães, transferem para o PB traços fonológicos característicos dessas línguas.

Alguns traços fonéticos característicos da variedade paulistana tradicional se devem ao contato de falantes de diferentes dialetos italianos com o PB. Por exemplo, a pronúncia desnasalizada das vogais tônicas de *homem, fome, Antônio, toma* (do verbo *tomar*), *viemos, demos* que soam [ɔ] e [ɛ] respectivamente, enquanto a maioria dos brasileiros as pronunciam como [õ] e [ẽ], bem nasalizadas. Isso se deve à inexistência de vogais nasais em italiano. Outra característica paulistana é o não alçamento das vogais [o] e [e] em sílabas pretônicas, alçamento que ocorre numa ampla porção do território brasileiro: as palavras *borracha, bolacha, fogão, futebol, advogado, pneu* soam [bo'haʃa], [bo'laʃa], [fo'gɐ̃ʊ], [fute'bɔʊ], [adevo'gadu], [pe'neʊ], enquanto em outras variedades soam [bu'haʃa], [bu'laʃa], [fu'gɐ̃ʊ], [fuʧi'bɔʊ], [adʒivo'gadu], [pi'neʊ], com as vogais mais fechadas e altas no lugar das menos fechadas e médias. Decerto, tendo de aprender o PB como nova língua e baseando-se na forma escrita das palavras, os falantes de dialetos italianos mantiveram as vogais menos fechadas e médias. Também é típica da variedade paulistana a pronúncia de palavras como *fazenda, entendendo* e *documento* como [fa'zɛ̃ẽda], [ĩtẽ'dẽẽdu] e [doku'mẽẽtu] com uma semiditongação do [e] diante da nasal, fenômeno que não ocorre nas demais variedades do PB, onde o que se tem é sempre a vogal nasal [ẽ].

As mudanças linguísticas ocorridas, portanto, nas línguas românicas ocidentais seriam o resultado do contato do latim com as línguas célticas, contato provocado pela *conquista* romana e que provocou, por sua vez, o desaparecimento total das línguas célticas continentais [▶208]. O latim constitui, desse modo, o *estrato* principal das línguas derivadas dele, enquanto as línguas célticas constituíram o *substrato*, no caso das românicas ocidentais.

Convém, por fim, assinalar que a hipótese da influência do substrato não encontra unanimidade entre os linguistas. Ela faz projeções sobre um passado muito remoto, do qual não se tem nenhuma evidência concreta. No caso que analisamos aqui, por exemplo, não se tem conhecimento preciso de como eram de fato as línguas faladas na Galécia antes da conquista romana. Por isso é sempre bom empregar essa hipótese com cuidado e discernimento, sem fazer afirmações muito categóricas.

4.3.3.3 Causas do contato: a colonização/escravização

Outro fenômeno social capaz de disparar a mudança linguística num ritmo acelerado e com consequências bastante radicais é a colonização/escravização. Logo em seguida às primeiras grandes navegações que descortinaram o resto do mundo para os europeus, as potências coloniais, repetindo em escala muitíssimo maior a formação do antigo Império Romano, se apoderaram das terras "descobertas" (invadidas, na verdade) e não hesitaram em submeter os povos conquistados ao massacre sistemático (como no caso dos indígenas de todo o continente americano) e/ou à escravização (como também ocorreu com os indígenas e, após sua quase total aniquilação, com os negros africanos). Na escravização dos negros africanos, Portugal teve papel pioneiro e preponderante.

A monstruosidade do tratamento dispensado pelos espanhóis aos índios americanos foi registrada com horror pelo religioso Bartolomé de Las Casas, em obra escrita em 1552, ou seja, pouco mais de meio século depois da chegada de Colombo no continente:

> Nessas ovelhas mansas [os índios] [...] entraram os espanhóis, tão logo as conheceram, como lobos e tigres e leões crudelíssimos famintos de muitos dias. E outra coisa não têm feito nos últimos quarenta anos, até hoje, e no dia de hoje o fazem, senão despedaçá-las, matá-las, angustiá-las, afligi-las, atormentá-las e destruí-las pelas estranhas, novas, várias e nunca lidas nem ouvidas maneiras de crueldade [...] em tal grau que, havendo na ilha Hispaniola [*atuais Haiti e República Dominicana*] mais de três milhões de almas que vimos, hoje não há dos naturais dela duzentas pessoas. A ilha de Cuba é quase tão longa quanto de Valladolid [*cidade da Espanha*] a Roma; está hoje quase toda despovoada. [...] Daremos por conta muito certa e verdadeira que morreram nesses quarenta anos pelas citadas tiranias e infernais obras dos cristãos, injusta e tiranicamente, mais de doze milhões de almas, homens e mulheres e crianças; e creio, na verdade, sem julgar enganar-me, que são mais de quinze milhões.

Com a virtual extinção dos povos indígenas de muitas áreas conquistadas, os europeus se voltaram para a exploração da mão de obra dos negros africanos, ini-

ciando um comércio de seres humanos que se estenderia por quase quatrocentos anos [▸235].

Antes de iniciar o sequestro e transporte forçado de africanos para o continente americano, os portugueses já mantinham relações comerciais com populações da África, em áreas que se tornariam colônias do império português. Dessas relações comerciais e de seus consequentes *contatos linguísticos* surgiram as chamadas **línguas crioulas**.

O processo de surgimento dessas línguas se dá em diferentes etapas. Inicialmente, para conseguir se comunicar com o europeu, a população local desenvolve um código rudimentar baseado na língua do estrangeiro. Esse código é chamado de pidgin. Com o passar do tempo e com a intensificação das relações, o *pidgin* se generaliza entre a população local. Numa nova etapa, esse meio de comunicação surgido do contato acaba por se tornar língua materna dos residentes, transmitida de geração em geração. É quando se constitui um **crioulo**. Evidentemente, para poder desempenhar todas as funções sociais, culturais, identitárias de uma língua humana, o crioulo se torna um sistema linguístico completo e complexo, com fonologia própria, gramática própria, léxico próprio. Um crioulo é classificado como *de base X* de acordo com a língua europeia da qual se originou: existem crioulos de base portuguesa, inglesa, francesa, espanhola e holandesa, resultantes do processo colonial-imperial moderno.

As chamadas **línguas crioulas** são, portanto, **línguas como quaisquer outras** que não recebem esse nome. Imaginar que sejam línguas "estropiadas", "rudimentares" ou "primitivas" é tão absurdo quanto julgar que o francês, o espanhol ou o PB são formas "estropiadas" ou "rudimentares" do latim. A única diferença entre as línguas crioulas e as outras línguas é sua origem no processo de colonização iniciado no final do século XV.

O linguista congolês Salikoko Mufwene, nascido no antigo Zaire e professor da Universidade de Chicago, tem se dedicado intensamente ao estudo das línguas crioulas. Para ele,

> os crioulos são especiais do ponto de vista epistemológico somente por um acidente do modo como nós temos feito linguística, não porque eles tenham se desenvolvido por qualquer processo evolutivo que não tenha ocorrido nos desenvolvimentos de outras línguas, nem porque sua gênese esteja inserida em ecologias sócio-históricas drasticamente diferentes do tipo em que evoluíram as línguas não crioulas, nem sequer porque representem algum tipo estrutural global de sistemas linguísticos. São tão naturais quanto as línguas não crioulas. (Mufwene, 2001: xi)

Para Mufwene, os crioulos em nada se diferenciam de outros processos de reestruturação linguística disparados pelo contato entre línguas, como ocorreu, por exemplo, na formação das línguas românicas a partir do latim. Afinal, as línguas românicas são tão diferentes do latim quanto o cabo-verdiano e o forro são diferentes do português, quanto o haitiano do francês, ou o tok-pisin do inglês. O pesquisador afirma que muitas das

supostas especificidades das línguas crioulas são, de fato, ilusões teóricas criadas pelo fato de que os crioulos surgiram do processo colonial e são falados predominantemente por negros. Ele acusa, sem rodeios, a ciência linguística de se deixar guiar por critérios de *etnicidade*, ou seja, por um óbvio racismo, ainda que não explícito.

O contato linguístico é um elemento acelerador dos ritmos da mudança linguística. Por isso as línguas crioulas aparentam ser transformações radicais das gramáticas das quais se originaram. No entanto, transformações igualmente radicais se processaram na formação histórica da língua francesa — decerto a que mais se distancia do latim — e da língua inglesa — muitíssimo diferente do anglo-saxão falado na Inglaterra após a invasão da ilha por populações vindas da Alemanha e da Escandinávia.

Para o linguista francês Bernard Cerquiglini, o francês é "um crioulo que deu certo", isto é, uma língua surgida de transformações muito profundas e que ganhou importância política e se impôs como língua nacional e imperial. Aspectos gramaticais geralmente atribuídos às línguas crioulas estão perfeitamente representados no francês padrão moderno, como a eliminação das marcas morfológicas de plural nos substantivos e adjetivos e a indicação de pluralidade somente no elemento determinante (artigo, possessivo, demonstrativo etc.). Assim, *la belle fleur rouge* ('a bela flor vermelha') se escreve no plural *les belles fleurs rouges* ('as belas flores vermelhas') — apesar das terminações em *-s* de todas as palavras no plural, esse *s* jamais é pronunciado e representa simplesmente a sobrevivência na escrita de algo que deve ter sido pronunciado muitos séculos atrás. A transcrição fonética deixa evidente que a única diferença entre as duas frases é a pronúncia do artigo definido, que de *la* passa a *les*:

> la belle fleur rouge → [la bɛl flœr ruʒ]
> les belles fleurs rouges → [le bɛl flœr ruʒ]

É o mesmo fenômeno que se observa no PB, tanto em suas variedades "populares" quanto na fala urbana culta distensa, em que a marca de plural fica restrita ao primeiro elemento do sintagma: *aquelas menina bonitinha*.

Outro aspecto sempre enfatizado no estudo dos crioulos é a simplificação dos paradigmas verbais, com o pronome-sujeito sempre explícito e o verbo numa forma única para todas as pessoas. Ora, essa é uma descrição que se encaixa perfeitamente no inglês. Nessa língua, como se sabe, os verbos regulares apresentam uma única forma para todas as pessoas no presente, com exceção da não-pessoa, que recebe uma terminação *-s* no singular: *I, you, we, they love; he/she/it loves*. Já na conjugação do passado, a forma verbal é sempre a mesma para todas as pessoas: *loved*. Aqui também é possível comparar o que ocorre em inglês com o que se verifica nas variedades do PB em que o verbo se organiza de forma semelhante: *eu, tu, você, ele/ela, nós, a gente, vocês, eles/elas amava*.

O processo de formação das línguas crioulas é idêntico ao da formação das línguas que não recebem essa classificação. É frequente a referência à *simplificação gramati-*

cal operada nas línguas crioulas quando comparadas às suas línguas-bases. Essa simplificação, no entanto, é resultante dos processamentos sociocognitivos que vamos examinar mais abaixo, nos quais opera a chamada *economia linguística*. O estudo histórico de toda e qualquer língua revela essas simplificações: perda da morfologia de casos, redução das flexões, fixação de uma ordem dos constituintes da sentença, eliminação de redundâncias etc. No inglês e nas línguas românicas, por exemplo, desapareceu a morfologia dos casos, restando apenas alguns vestígios (o genitivo em inglês, com o uso de *'s — this is Peter's house* ['essa é a casa do Pedro']; a variação morfológica dos pronomes segundo a função sintática: *eu*-sujeito, *me*-objeto). Em alemão, o paradigma dos casos vem se simplificando: o acusativo, por exemplo, só é marcado nas palavras do gênero masculino. Os três gêneros do latim (feminino, masculino e neutro) se reduziram a dois nas línguas românicas (feminino e masculino). Em inglês, os gêneros desapareceram por completo nos substantivos, adjetivos e determinantes. O número e o tipo de tempos verbais é muito diferente em latim e nas línguas românicas, que criaram novos tempos e abandonaram distinções que existiam em latim.

As simplificações ocorridas na formação das línguas crioulas são atribuídas à forma imperfeita, irregular, assistemática com que os falantes locais aprenderam a língua do colonizador. No entanto, cabe perguntar se não foi isso também o que ocorreu, por exemplo, quando os celtas que habitavam a Península Ibérica passaram a usar o latim. Além disso, assim como as línguas crioulas se formaram com base em variedades faladas muito distantes das normas literárias — já que os marinheiros, mercadores e colonos não pertenciam à reduzidíssima elite letrada de seus países de origem —, também as línguas românicas se formaram com base no chamado *latim vulgar*, que nem de longe se assemelhava à língua dos grandes oradores da Roma clássica. É portanto de um conjunto de variedades próprias das camadas menos prestigiadas da população e de um aprendizado assistemático que as línguas surgidas de contato se formaram e se formam.

Nos territórios onde surgiram as línguas crioulas, é comum ocorrer um processo de reaproximação da língua surgida do contato com a língua que lhe serviu de base. Esse processo costuma ser designado como *descrioulização*, um termo muito criticado por alguns linguistas. No caso dos crioulos de base portuguesa, por exemplo, costuma se falar de *reaportuguesamento* da língua. Nesse processo, algumas características próprias das línguas crioulas, como por exemplo a ausência de marcas de gênero ou o paradigma verbal reestruturado, estariam sendo abandonadas em proveito de formas características da língua-base, que sempre goza de maior prestígio. Dado que as línguas chamadas crioulas não contam com *status* oficial, elas também não contam com uma ortografia unificada nem com uma norma linguística elaborada para servir de código comum a todos os falantes. Com isso, essas línguas se encontram em ampla variação, uma vez que não sofrem processos centrípetos de regulação. Em Cabo Verde, por exemplo, cada uma das nove ilhas habitadas do arquipélago tem sua variedade local de cabo-verdiano, com algumas diferenças bem marcantes. Sendo o português a língua oficial do ensino e da administração pública, não espanta que sua influência sobre a língua mais falada seja tão grande.

No entanto, convém lembrar que fenômenos semelhantes ocorreram com as grandes línguas europeias. No período conhecido como Renascimento, muitas línguas nacionais europeias se tornaram oficiais, substituindo o latim como língua do Estado. Com isso, essas línguas passaram por um processo de normatização: é precisamente nessa época que surgem as primeiras gramáticas modernas, sendo a primeira de autoria de Antonio de Nebrija, que publicou sua *Gramática de la lengua castellana* em 1492. Nesse processo de elaboração de uma norma linguística que sirva para a administração pública, o ensino, a discussão filosófica, a produção literária etc., os codificadores das línguas recém-oficializadas se voltaram para o latim clássico, fazendo amplo uso do léxico e da gramática da língua mais prestigiada durante dezenas de séculos. Justamente por isso esse período é chamado de relatinização. Nesse período, muitos termos técnicos, científicos, filosóficos e literários foram criados tendo por base o étimo latino. Esses termos são produto do que se chama *formação erudita*, em contraste com a *formação popular*, na qual os processos de mudança linguística ocorreram regularmente. Muitos desses termos eruditos, por exemplo, são proparoxítonos, uma acentuação regularmente eliminada na história da língua em favor das palavras paroxítonas. Observe no quadro abaixo os pares de palavras de formação erudita e de formação popular: a datação mostra quando elas começaram a ser usadas e revela que as palavras eruditas são muito mais recentes na história da língua do que as de formação popular. Além da relatinização do léxico e da gramática, com a reintrodução de construções sintáticas próprias do latim clássico, também se deu um processo de *recondução* de diversas palavras a uma forma mais próxima do étimo latino. Por exemplo, a palavra *letradura* foi substituída por *literatura*, assim como *esmar* por *estimar*, *frol* por *flor*, *marteiro* por *martírio*, *chantar* por *plantar*, *coa* por *cauda*, *seenço* por *silêncio* etc.

LATIM	PORTUGUÊS			
	FORMAÇÃO ERUDITA	DATAÇÃO	FORMAÇÃO POPULAR	DATAÇÃO
dígitu-	dígito	1532	dedo	séc. XIII
duplu-	duplo	1651	dobro	1277
frígidu-	frígido	1542	frio	1101
mácula-	mácula	1589	malha	séc. XIII
másculu-	másculo	1687	macho	séc. XIII
matéria-	matéria	séc. XIV	madeira	1269
óculu-	óculo	1649	olho	séc. XIII
plaga-	plaga	séc. XIV	chaga	1297
planu-	plano	séc. XIV	chão	1261
rígidu-	rígido	1572	rijo	séc. XIII

O chamado processo de "descrioulização", portanto, guardadas as devidas diferenças socioculturais e políticas, se assemelha ao que ocorreu com as línguas europeias no Renascimento, quando foram oficializadas e normatizadas. É mais do que provável que, se alguma língua "crioula" vier a se tornar oficial, os processos de normatização e codificação pelos quais ela passará implicarão no recurso à herança lexical e morfossintática da língua da qual ela se originou.

Tudo isso tem levado alguns teóricos, como Mufwene e Cerquiglini, a afirmar que, na prática, todas as línguas do mundo poderiam ser chamadas de crioulas, uma vez que é raríssimo encontrar línguas isoladas cujos falantes nunca tenham entrado em contato com outras línguas e nas quais não tenham ocorrido os processos de mudança inerentes a toda e qualquer língua humana. Assim, ou todas as línguas são crioulas, ou nenhuma língua é.

4.3.3.4 O contato linguístico na formação do português brasileiro

No capítulo 5, vamos nos dedicar ao estudo da formação histórica do PB e da importância do elemento africano na configuração atual da nossa língua. Cada vez mais autores reconhecem que as diferenças marcantes entre o PB e a língua da qual se originou — o português europeu em sua fase de transição do período medieval para o moderno — se devem primordialmente ao multilinguismo que caracterizou a história do Brasil na maior parte do período colonial. A dispersão pelo território brasileiro de milhões de negros escravizados, falantes de muitas línguas diferentes, não pode ter deixado de incidir fortemente sobre o desenvolvimento do PB. Durante mais de três séculos, a população negra e mestiça representou a maioria dos habitantes do Brasil: na época da Independência, 75% da população era composta de não brancos.

Uma questão que sempre despertou o interesse dos pesquisadores é por que não surgiu no Brasil, apesar do intenso afluxo de escravos africanos, uma língua nova, "crioula", radicalmente distinta do português europeu, como o cabo-verdiano, o forro, o *cristang* etc. As semelhanças entre o PB e o português europeu são mais nítidas do que entre essas duas línguas e as demais surgidas do processo de colonização. Dante Lucchesi, um dos mais destacados pesquisadores brasileiros dedicados ao estudo do influxo africano na formação da nossa língua, oferece as seguintes respostas (Lucchesi, 2009: 70):

- **a proporção entre a população de origem africana e branca, que permitiu um maior acesso à língua-alvo [o português] do que o observado nas situações típicas de crioulização:**

"[...] a proporção de população branca no Brasil nunca foi inferior a 30%, crescendo significativamente no século XIX, quando chega a quase metade do total. Esse quadro está bastante distante do que deu ensejo às formas mais típicas de crioulização, ocorridas no Caribe. No Haiti e na Jamaica, a proporção de brancos nunca ultrapassou os 10% durante praticamente todo o período da colonização, e o nível de miscigenação entre brancos e negros foi muito mais baixo do que o observado no Brasil. Portanto, de um modo geral, o acesso aos modelos da língua-alvo no Brasil sempre foi maior do que o que se observa nas situações mais típicas de crioulização. (p. 62-63)

- **a ausência de vida social e familiar entre as populações de escravos, provocada pelas condições sub-humanas de sua exploração, pela alta taxa de mortalidade e pelos sucessivos deslocamentos:**

A crioulização depende de forma crucial da nativização da língua-alvo, o que, por sua vez, passa pela socialização do uso dessa língua-alvo entre a população dominada. [...] A situação de vida da maioria dos escravos brasileiros não lhes permitia o acesso a qualquer forma de vida familiar ou social. (p. 63)

- **o uso de línguas francas africanas como instrumento de interação dos escravos segregados e foragidos:**

A crioulização depende do preenchimento de lacunas que se dá quando a língua dominante [o português] é também usada na interação entre os escravos. Se, nessa interação, os escravos lançam mão de uma língua franca baseada na língua nativa do segmento étnico africano mais representativo e/ou de maior prestígio, o processo de crioulização é inibido (p. 64). No panorama geral dos três séculos de tráfico há um grande predomínio de escravos trazidos da zona linguística banto. [...] Esse predomínio banto, sobretudo nos séculos XVI e XVII, reflete-se na formação de línguas gerais africanas no Brasil [...] (p. 65)

- **o incentivo à proficiência em português:**

[...] outros fatores como a integração do negro na sociedade brasileira e, sobretudo, a miscigenação também atuaram contra a formação de uma língua crioula no Brasil. Se a maioria dos escravos se destinava ao trabalho da lavoura, onde a aquisição do português se dava de forma bastante precária, há de levar-se em conta também os escravos que eram destinados aos serviços urbanos e aos trabalhos domésticos. Esses provavelmente tinham um maior acesso aos modelos da língua-alvo e adquiriam uma variedade de português mais próxima da língua de seus senhores. Para além disso, deve ser considerado que a maior proficiência em português sempre trouxe mais vantagens aos escravos, mesmo entre os que se encontravam nos engenhos [...] (p. 67)

- **a miscigenação racial:**

Por fim, o fator que foi decisivo para que não tenha ocorrido um processo representativo de crioulização do português no Brasil foi a miscigenação racial. [...] Alargando os níveis de interação sociocultural entre os setores dominantes e dominados, a miscigenação foi descortinando progressivamente, para os indivíduos mestiços, novas vias que lhes permitiam uma maior integração na sociedade. [...] Portanto, a miscigenação impediu que, no Brasil, a segregação racial confinasse a população de africanos e seus descendentes em guetos sociais de um mundo cultural à parte, do qual a formação de uma língua crioula e claramente diferenciada [...] seria a consequência natural. (p. 68-69)

De fato, quando comparamos a situação dos negros escravos no Brasil e, por exemplo, nos Estados Unidos, as diferenças ficam bastante claras. Nesse país, a profunda e duradoura

segregação racial, que permaneceu amparada pela legislação até a década de 1960, e a virtual inexistência de miscigenação — decorrente, entre outras coisas, da criminalização das relações inter-raciais por leis que vigoraram também até 1967 —, fez surgir uma língua característica dos guetos urbanos das grandes cidades americanas, língua que já foi denominada BEV (*Black English Vernacular*), atualmente designada na literatura sociolinguística como AAVE (*Afro-American Vernacular English*), e que foi batizada de *Ebonics* (derivado de *ebony*, 'ébano') por intelectuais negros na década de 1970.

Ao contrário dos Estados Unidos, não se pode dizer que no Brasil exista um "português dos negros": o que existe é uma polarização, decorrente da profunda segregação social que tem caracterizado a nossa sociedade, entre a língua dos segmentos mais pobres — a maioria da nossa população, composta de brancos e, mais essencialmente, de não brancos — e a língua dos segmentos mais ricos — essencialmente brancos. As variedades linguísticas mais estigmatizadas em nossa sociedade são faladas por negros, índios, mestiços e brancos com menor acesso à escolarização, ao trabalho e à renda. Embora, no plano social, o Brasil seja um país impregnado de racismo, no plano linguístico, as diferenças que separam as variedades urbanas privilegiadas das demais, estigmatizadas, são de ordem socioeconômica: a gramática dos negros pobres e dos brancos pobres é a mesma. É o grau de escolarização o principal fator social que diferencia as variedades mais prestigiadas das variedades mais estigmatizadas.

Os aspectos linguísticos mais estigmatizados pelos falantes urbanos cultos ocorrem nos modos de falar de negros, índios, mestiços e brancos pobres: a restrição das regras de concordância nominal, a simplificação do paradigma verbal, a rotacização de [l] em encontros consonantais (*pranta, crima, ingrês*) ou em travamento silábico (*fi[r]me, fa[r]ta, cu[r]pa*), a lambdacização, velarização ou vocalização de [r] em travamento de sílaba (*ga[ʊ]fo, te[ʊ]ça, ce[ʊ]veja*), a deslateralização da consoante [ʎ], que se vocaliza (*traba[ɪ]o, pa[ɪ]a, abe[ɪ]a*) etc.

Os aspectos mais característicos do PB, especialmente em suas variedades estigmatizadas rurais ou rurbanas, mas também presentes (ainda que com menor frequência) nas variedades urbanas prestigiadas, decorrem, portanto, do contato entre o português colonial e as línguas africanas trazidas com os escravos. Esse contato, como sempre ocorre, acelerou os fatores inerentes à mudança linguística (de ordem cognitiva), além de permitir que aspectos próprios às línguas africanas (especialmente as do grupo banto) atuassem como substrato na constituição do PB. Disso a gente vai se ocupar mais demoradamente no próximo capítulo.

4.4 Fatores sociocognitivos e mudança linguística

A mudança linguística é um processo *sociocultural e sociocognitivo*, ou seja, um processo que tem origem na interação entre a dinâmica social da comunidade de fala e o processamento da língua no cérebro por parte dos indivíduos em interação sociocomunicativa. São três as forças inerentes mais importantes na mudança linguística:

(1) a economia linguística;
(2) a gramaticalização;
(3) a analogia.

4.4.1 A *economia linguística*

Economia linguística é um termo que recobre uma variada gama de processos que se caracterizam por representar mecanismos de mudança que tentam reagir positivamente a dois impulsos: (a) poupar a memória, o processamento mental e a realização física da língua, eliminando os aspectos redundantes e as articulações mais exigentes; (b) preencher lacunas na gramática da língua, de modo a torná-la mais eficiente como instrumento de interação sociocomunicativa.

Como se vê, a economia linguística *elimina*, de um lado, mas também *cria* ou *recria*, do outro. Por isso, não deve ser identificada com a célebre "lei do menor esforço" postulada pelos linguistas do século XIX, em correspondência com a ideologia romântica que vigorava na época. A "lei do menor esforço" não se exercia sobre as línguas faladas por povos "primitivos" ou "selvagens", mas sim nas línguas faladas pelas pessoas que se tornaram "mal acostumadas" aos confortos e privilégios da vida urbana moderna, "civilizada". Compartilhando da conhecida idealização romântica do "bom selvagem", o linguista alemão Max Müller dizia que as pessoas habituadas à indolência da vida civilizada não usavam os movimentos articulatórios vigorosos exigidos pelas línguas dos povos "primitivos", por isso as línguas ocidentais, ao longo do tempo, iam perdendo cada vez mais coisas e cada vez se tornavam mais brandas, relaxadas etc. Essa visão, perfeitamente sintonizada com as crenças da época, em nada corresponde à concepção científica que se tem hoje do fenômeno.

É preferível, atualmente, postular que existem tendências que pressionam as línguas de maneira inevitável por causa da configuração articulatória, fisiológica e psicológica da espécie humana. Por exemplo, no que diz respeito à estrutura silábica das palavras nas diferentes línguas do mundo, as pesquisas revelam uma tendência ao padrão cvcv, isto é, consoante-vogal-consoante-vogal, como em *casa*, porque, do ponto de vista articulatório, é mais cômodo iniciar a palavra com uma oclusão ou uma fricção do que com uma vogal, que exige maior quantidade de ar, e é mais cômodo também terminar a palavra com uma vogal, porque não exige o esforço físico de pronunciar uma consoante em final de palavra. Isso explica, por exemplo, por que em francês, ao longo do tempo, as consoantes finais deixaram de ser pronunciadas, embora ainda venham escritas por força de um empedernido conservadorismo ortográfico:

- grand [grã] ('grande')
- fort [fɔr] ('forte')

- tout [tu] ('tudo', 'todo')
- voix [vʊa] ('voz')
- trop [tro] ('demais')
- mais [mɛ] ('mas')

Essa mesma tendência é claramente visível no PB, uma língua em que já são poucas as consoantes que podem ocorrer em final de palavra: [l], [r], [s], não por acaso consoantes que têm características bem especiais com relação às outras [▶333]. O [r] e o [l] são chamadas *soantes* (ou *líquidas*), porque em sua articulação se aproximam bastante das vogais, razão por que também é comum que se transformem em verdadeiros sons vocálicos, e em algumas línguas (como as eslavas) possam até constituir centro de sílaba: *smrt* ('morte' em servo-croata e tcheco), *Krk* (nome de uma ilha do litoral da Croácia), *Hrdlčka* (sobrenome de um conhecido antropólogo tcheco). Em inglês também é possível encontrar muitas palavras em que o [l] atua como uma vogal: *world* ('mundo'), *girl* ('menina'), *snarl* ('rosnar'), *gnarl* ('nó da madeira'), *earl* ('conde'), *battle* ('batalha'), *little* ('pequeno') etc. As soantes também incluem as nasais [m] e [n] que, em final de sílaba e, portanto, de palavra, há séculos se transformaram, em português, em mera reverberação nasal da vogal anterior, de modo que a lista de dígrafos da nossa língua também inclui *am, em, im, om, um, an, en, in, on, un*: *bombom* [bõ'bõ], *cantando* [kã'tãdu], *ninguém* [nĩ'gẽ̃]. Também não é por acaso que essas consoantes, mais o [s], são as únicas que podem travar sílabas na nossa língua: *campa, canta, carta, calma, casta*. A consoante [s] é classificada como *contínua*, porque sua emissão pode se prolongar indefinidamente, enquanto o falante tiver ar para tanto, o que já não é possível fazer com as oclusivas do tipo [b], [d], [g], [p], [t], [k], [f], [v] etc.

Mesmo sendo reduzido o número de consoantes que podem ocorrer em final de palavra — apenas três: [l], [r], [s] como vamos ver em [▶333] —, elas sofrem forte pressão para não ocupar esse lugar ou, então, para se transformar em vogais. O [l] na grande maioria das variedades do PB já se vocalizou completamente em [ʊ]: *mal* [maʊ], *mel* [mɛʊ], *mil* [miʊ], *gol* [goʊ], *sul* [suʊ].

Em outras variedades, menos difundidas e restritas a algumas zonas rurais, ocorre o fenômeno chamado *paragoge*, isto é, o acréscimo de uma vogal à consoante final da palavra: [sɔli] para *sol*; [pa'pɛli] para *papel*; [mali] para *mal* etc. A paragoge também é característica do português europeu quando a palavra termina em [r], de modo que os infinitivos verbais *cantar, beber, subir* naquela língua soam [kã'tarə], [be'βerə], [su'βirə] etc. No Brasil, o tratamento dado ao [r] de final de palavra se faz preferentemente por *apócope*, isto é, por queda da consoante: [kã'ta], [be'be], [su'bi], [profe'so] etc. Uma vez que a alternância da sílaba tônica (junto, em alguns casos, com a alternância de vogal alta e vogal baixa) já permite distinguir o infinitivo do verbo conjugado, o [r] pode ser dispensado para a identificação do infinitivo. Compare:

presente	infinitivo
['abri]	[a'bri]
['fala]	[fa'la]
['kãta]	[kã'ta]
['ovi]	[o'vi]
['sabi]	[sa'be]
['vivi]	[vi've]

Quanto ao [s] final, que ocorre sobretudo na sua forma de morfema de plural, sua supressão é muito frequente nos sintagmas nominais, em que a marca de plural se fixa no determinante, de modo que o [s] se torna de fato uma consoante de final de sílaba numa dupla de palavras que constituem uma só: *as casa* [as'kaza], *as outra* [a'zotra], *os prato* [us'pratu], *os home* [u'zõmi] etc. Essa economia articulatória corresponde também a uma economia de recursos morfológicos: a eliminação das marcas redundantes, como vamos ver logo adiante. Convém lembrar que a regra de concordância <marque o plural somente no determinante> ocorre na fala de todos os brasileiros, independente de sua origem rural ou urbana, mais ou menos escolarizados etc. O que distingue o uso de mais ou menos marcas de concordância é a **frequência de sua realização**: os falantes urbanos mais letrados tendem a fazer mais concordância e em situações comunicativas mais monitoradas; mesmo assim, essa concordância nunca chega a 100% dos casos [▶994]. Outro caso de supressão do [s] é na flexão verbal da 1ª pessoa do plural: mesmo os falantes urbanos letrados, em situações de interação mais distensas, tendem a suprimir o [s] da terminação *-mos*, de modo que pronunciam [fi'zẽmu], [kã'tãmu], [ka'mu], para o que se escreve *fizemos, cantamos, caímos*. Muito comum também é a forma ['vãmu], *vamos*, para a formação do imperativo na 1ª pessoa do plural no PB [▶569]. Por fim, vamos lembrar as variedades rurais e rurbanas em que o [s], mesmo não sendo marca de plural, desaparece após a inserção de uma semivogal [ɪ]: *luz* [luɪ], *mês* [meɪ], *vez* [veɪ] etc.

O [s] em sílaba travada sofreu eliminação radical na história do francês. No interior das palavras, ele passou por uma primeira fase de aspiração antes de desaparecer por completo:

- feste > fe[h]te > fête ('festa')
- hospital > ho[h]pital > hôpital ('hospital')
- mesme > me[h]me > même ('mesmo')

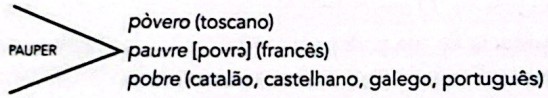
Na escrita do francês, se usa um acento circunflexo para indicar a perda desse [s], que deixou de ser pronunciado séculos atrás. É só mais uma das complicações injustificáveis da ortografia dessa língua. O *[s]* como marca de morfema plural também desapareceu em francês, embora se conserve na escrita. Diferentemente do PB, no entanto, em que o [s] é bem pronunciado no determinante (*as casa*), em francês até essa marca de plural desapareceu, de modo que a diferença entre singular e plural

se faz exclusivamente pelo timbre da vogal: *le livre* [lə'livrə]/*les livres* [le'livrə] ('o livro'/'os livro[s]'); *ce livre* [sə'livrə]/*ces livres* [se'livrə] ('esse livro'/'esses livro[s]'). Na grande maioria dos verbos, não existe nenhuma diferença fonética entre a não-pessoa do singular e a não-pessoa do plural, de modo que somente o contexto de enunciação pode revelar o número: *il parle* [il'parlə]/*ils parlent* [il'parlə] ('ele fala'/'eles fala[m]').

Em muitas línguas do grupo espanhol, o [s] em travamento de sílaba, em final de palavra ou como morfema plural é pronunciado como uma simples aspiração — *las cosas* [lah'kosah], *los hombres* [loh'ombreh] — ou simplesmente desaparece no nome [lah'kosa], [loh'ombre], conservando-se apenas no determinante.

São muitos os processos conhecidos de economia linguística no plano articulatório, vários deles muito importantes na história das línguas. Vamos citar aqui apenas alguns para ilustração:

1. ASSIMILAÇÃO: atração de um determinado som sobre outro, de modo a torná-lo semelhante ao primeiro. Um caso frequentíssimo é o da transformação do ditongo latino [aʊ] em [o] nas diferentes línguas românicas:



(a) Nesse processo, a semivogal [ʊ], alta e fechada, "puxou" para si a vogal [a] baixa e aberta. A grafia de muitas palavras em português revela um estágio anterior em que [aʊ] primeiro se transformou em [oʊ]: *paucu- > pouco; lauru- > louro; raucu- > rouco* etc. Apesar disso, como se sabe, esse ditongo há muito se monotongou, tanto no PB quanto no português europeu. No galego, mais conservador nesse aspecto, o ditongo escrito *ou* ainda se pronuncia [oʊ]. A assimilação teria sido total se [aʊ] tivesse se transformado em [u]. No entanto, ela ficou a meio caminho entre o [ʊ] muito alto e o [a] muito baixo: [o]. Em francês, no entanto, ocorreu uma assimilação total do ditongo escrito *ou*, que se pronuncia sempre como [u]: *pou* [pu] ('piolho'), *tout* [tu] ('tudo'), *coup* [ku] ('golpe', 'corte'), *doux* [du] ('doce') etc.

(b) Evolução semelhante ocorreu com o ditongo [aɪ], que não existia em latim. Nas palavras latinas terminadas em *-ariu*, por exemplo, se processou uma *metátese*, isto é, a troca de lugar do *i* com o *r*: *-airu*, também na busca de uma economia de articulação (de três sílabas em latim clássico — *a-ri-u* — para duas no latim vulgar e nos romances: *ai-ru*). Logo em seguida, por força de atração do *i*, o *a* se fechou em *e*:

ianuariu- > janairu- > janeiro
primariu- > primairu- > primeiro
solitariu- > solitairu- > solteiro

A partir dessa última fase, muito remota, as diferentes línguas do grupo portugalego deram tratamentos diferenciados ao ditongo [eɪ]. Na variedade lisboeta, ocorreu uma *dissimilação* (ver logo abaixo). No PB, diante de determinadas consoantes (palatais e vibrante), o ditongo se monotongou em [e]. No galego, a pronúncia mais antiga [eɪ] permanece. No castelhano, a assimilação foi total, de modo que a ortografia só registra a terminação *-ero* (*primero, soltero, enero, obrero, brasilero, extranjero, dinero* etc.)

Em francês, a assimilação de [aɪ] foi total, uma vez que nessa língua esse ditongo se transformou na vogal [ɛ] ou [e], dependendo da variedade regional:

craie [krɛ] ('giz')
j'aimais [ʒɛmɛ] ('eu amava')
lait [lɛ] ('leite')
mais [mɛ] ('mas')
paix [pɛ] ('paz')

Em inglês, o grupo escrito *ay* sempre se pronuncia [eɪ]:

day [deɪ] ('dia')
gay [geɪ] ('alegre')
May [meɪ] ('maio')
pay [peɪ] ('pagar')
say [seɪ] ('dizer')

No inglês australiano, porém, a pronúncia [aɪ], mais antiga, se conserva, de modo que *today* ('hoje') e *to die* ('morrer') são pronunciados [tə'daɪ].

(c) Em inglês, o grupo escrito *-stl-* ou *-st-* é pronunciado sem o [t], que foi assimilado totalmente pelo [s]: *castle* ['kəsəl] ('castelo'), *whistle* ['ʊisəl] ('assobio'), *listen* ['lisn] ('ouvir') etc. No PB, esse mesmo fenômeno ocorreu com o demonstrativo *este*, que desapareceu da língua falada, tendo se assimilado a *esse*. Em algumas variedades regionais onde se usa o pronome *tu*, as formas verbais terminadas em *-ste* se transformaram em *-se*: *tu foste* ['fosi]; *viste?* ['visi] etc.

(d) Assimilação semelhante à descrita acima ocorreu na história do português:

persona > persoa > pessoa
persicu- > pêssego
nostru- > nosso
ipse > esse

O galego conserva até hoje a forma *persoa*. No PB esse processo continua: em muitas variedades o grupo *-rs-* tende a ser pronunciado como *-s-* em palavras como *ma[s]o, unive[s]idade, comé[s]io* etc.

(e) No PB falado mais espontâneo, inclusive por pessoas altamente letradas de determinadas regiões (Minas Gerais, por exemplo), a terminação *-ndo*, característica do gerúndio, se reduziu a *-no*: *falano, correno, comeno, dormino* etc. Ocorreu aqui a assimilação do [d] pelo [n] subsequente. São duas consoantes dentais, de articulação muito próxima, facilmente assimiláveis. O *-ndo* deve ter passado por um primeiro estágio *-nno*, logo simplificado em *-no*. Assimilação idêntica ocorreu no napolitano, em que o gerúndio se faz em *-nno*. Essa mesma explicação serve para pronúncias comuns com *também* por *também*: assim como [n] e [d] têm articulação próxima, [m] e [b] são bilabiais, muitos predispostas à assimilação. Observe-se que a palavra *rumo* provém de um espanhol *rumbo*. Por outro lado, em espanhol existe *lamer* e *lomo*, para o que nós dizemos *lamber* e *lombo*.

2. DISSIMILAÇÃO: é o oposto da assimilação e resulta na *distinção* de dois sons muito próximos, para que não se confundam. Um caso conhecido de dissimilação é o do ditongo escrito *ei* no português europeu (variedade lisboeta), em que a vogal [e] é pronunciada [ɐ] diante da semivogal palatal, gerando o ditongo [ɐɪ], desconhecido em PB, em palavras como *leite, rei, farei, sei*, que para nossos ouvidos soam como "lâite", "râi", "farâi", "sâi" etc. Nos demais dialetos do português europeu, assim como em todo o PB, ocorre, ao contrário, assimilação e monotongação de [eɪ] diante de consoantes palatais e da vibrante simples [▶320].

(a) O que ocorre na variedade lisboeta também se verifica em alemão, onde o que se escreve *ei* é sempre pronunciado [aɪ]: *Einstein* ['aɪnʃtaɪn], *drei* [dʀaɪ] ('três'), *nein* [naɪn] ('não') etc. Outro caso de dissimilação ocorrido em alemão é o do ditongo *eu* que se mantém na escrita mas se pronuncia [ɔɪ]: *Freude* ['fʀɔɪdə] ('alegria'); *Feuer* ['fɔɪɐ̯] ('fogo'); *neu* [nɔɪ] ('novo') etc.

(b) Em inglês, o ditongo escrito *ow* ou *ou* muitas vezes é pronunciado [aʊ], de modo a deixar bem nítidos os dois sons envolvidos:

allow [ə'laʊ] ('permitir')
cow [kaʊ] ('vaca')
how [haʊ] ('como')
loud [laʊd] ('ruidoso')
now [naʊ] ('agora')
proud [pɹaʊd] ('orgulhoso')

A percepção dessa dissimilação leva muitos brasileiros a pronunciar a palavra *country* como ['kaʊntri], quando de fato a pronúncia dos falantes de inglês é ['kəntri], sem ditongo.

(c) Em francês, o ditongo escrito *oi* se pronuncia atualmente [ʊa]. No entanto, a forma escrita revela que, séculos atrás, esse ditongo era pronunciado [oɪ]. De fato, aqui a dissimilação agiu em duas etapas: para distinguir o [o] e o [ɪ], ambas fechadas, os falantes primeiro passaram a pronunciar o ditongo como [ʊɛ] e, mais tarde, para enfatizar ainda mais a distinção, como o atual [ʊa]. A passagem de [ʊɛ] para [ʊa] deve ter ocorrido no século XVIII, depois da chegada dos franceses no Quebec (Canadá), porque o francês canadense até hoje, em suas variedades não monitoradas, apresenta a pronúncia [ʊɛ] em palavras como *moi* ('mim'), *roi* ('rei'), *loi* ('lei') [mʊɛ, ʀʊɛ, lʊɛ], que os franceses pronunciam [mʊa, ʀʊa, lʊa]. Outro testemunho é o nome que damos ao instrumento de sopro chamado *oboé*, que entrou no português no final do século XVIII, proveniente do francês *haut-bois* ('madeira alta'), que naquela época devia ser ainda pronunciado como [o'bʊɛ] e não [o'bʊa], como dizem os franceses hoje.

3. EPÊNTESE: é a inserção de um segmento sonoro no interior de uma palavra, de modo a facilitar a articulação. No PB ocorre epêntese sistemática de uma vogal depois das chamadas consoantes 'mudas': *pneu* [pi'neʊ]/['peneʊ]; *advogado* [adivo'gadu]/[adʒivogadu]/[adevo'gadu]; *optar* [opi'tah] etc. Em algumas variedades, é comum a epêntese que rompe encontros consonantais, destrava sílabas travadas ou elimina hiatos: *trem* [te'ʀẽj], *parceiro* [pari'seru], *muamba* [mu'vãba] etc. O latim *blatta* nos deu *barata*, e *planu-* nos deu *porão*. A epêntese que elimina hiatos pela inserção de uma semivogal que facilita a transição entre as duas sílabas foi extremamente produtiva na história da nossa língua:

arena > *arẽa* > *area* > *areia*
foedu- > *feo* > *feio*
plenu- > *chẽo* > *cheo* > *cheio*
senu- > *sẽo* > *seo* > *seio*
vinu- > *vĩo* > *vinho*

(a) A epêntese é muito frequente no PB para a eliminação de hiatos em muitas variedades: *boa* [boʊa], *coroa* [koroʊa], à toa [atoʊa] etc. No português europeu, é comum a epêntese de uma semivogal [ɪ] para evitar o hiato entre o artigo feminino e a palavra seguinte, se essa começa com [a]: *a água* [ɐɪagʊɛ], fenômeno desconhecido do PB.

(b) É frequentíssima a epêntese do [ɪ] no PB entre uma vogal e um [s] em final de palavra, pronúncia que não sofre estigmatização: ao contrário, é tão corrente que os falantes das variedades que não a fazem são logo reconhecidos como provenientes de determinadas regiões (Minas Gerais, Rio Grande do Sul, por exemplo):

arroz [a'hoɪs]
dez [deɪs]
fiz [fiɪs]
mas [maɪs]
nós [noɪs]
pôs [poɪs]
voz [voɪs]

Essa epêntese explica algumas hesitações de muitos brasileiros na hora de escrever *mas*, uma vez que a pronúncia da conjunção para muitos deles é igual à do advérbio *mais*. Sendo um fenômeno tão generalizado, não existe justificativa para a grafia estereotipada "nóis", que muitas pessoas usam quando querem retratar a fala de pessoas supostamente incultas ou da zona rural. A epêntese nesse caso ocorre na fala de milhões de pessoas altamente escolarizadas, e só um forte preconceito social, aliado a um conhecimento superficial da variação linguística, explica a persistência dessa grafia.

Na fala de cariocas jovens, essa epêntese se dá também em sílaba travada por [s], caracterizando um traço geracional:

asma ['aɪʒma]
castigo [kaɪʃ'tigu]
festa ['fɛɪʃta]
honesto [on'ɛɪʃtu]
mostra ['mɔɪʃtra]

4. SÍNCOPE: é a eliminação de algum segmento fônico no interior de uma palavra. Quando a eliminação se dá no início da palavra, ela se chama *aférese* (*'brigado*, por *obrigado*); quando se dá no final da palavra, *apócope* (*batê* por *bater*).

(a) Um processo de síncope muito importante nas línguas românicas ocidentais foi a eliminação das vogais ou sílabas postônicas das palavras proparoxítonas. O latim não tinha palavras oxítonas, somente paroxítonas e proparoxítonas. Como depois da sílaba tônica havia ainda duas outras sílabas, o esforço articulatório para pronunciá-las foi aos poucos diminuindo, levando ao desaparecimento da sílaba imediatamente seguinte à tônica. Por isso, as palavras proparoxítonas que sobrevivem nas línguas ibéricas são praticamente todas palavras de formação erudita, isto é, termos científicos e técnicos, conceitos filosóficos, termos da linguagem poética, diretamente formados sobre o étimo latino. É o que explica a existência de pares como *coalho/coágulo, dedo/dígito, macho/másculo, rijo/rígido* etc. É o que se chama em linguística histórica de *formas divergentes*: duas palavras provenientes de um mesmo étimo. São incontáveis as palavras que em latim eram proparoxítonas e em castelhano, galego e português se tornaram paroxítonas devido à síncope da primeira vogal e/ou primeira sílaba postônica. O acento paroxítono é o da esmagadora maioria das palavras dessas línguas. Alguns exemplos:

ánima > *alma*
apícula > *abelha*
ásinu- > *asno*
cálidu- > *caldo*
clavícula > *chavelha*
cómputu > *conto*

genúculu- > *joelho*
ínsula > *ilha*
mácula > *mágoa* / *mancha* / *malha*
óculu- > *olho*
pópulu- > *povo*
régula > *régua* / *regra* / *relha*
vétulu- > *velho*

5. DESNASALIZAÇÃO DAS VOGAIS E DITONGOS NASAIS POSTÔNICOS E MONOTONGAÇÃO DESSES DITONGOS: muito cedo no latim vulgar as consoantes nasais em final de palavra sofreram apócope, ou seja, desapareceram. Na evolução das línguas românicas ocidentais, o léxico foi formado com base na forma do acusativo latino (caso do objeto direto), cuja terminação típica, no singular, é *-m*. Esse *-m* final desapareceu nas línguas românicas ocidentais, de modo que não é certo dizer que o latim *lupum*, por exemplo, é a fonte do castelhano, galego e português *lobo*, mas sim a forma *lupu-*, já sem a marca do acusativo. A pronúncia de uma nasalidade depois da sílaba tônica exige esforço articulatório, por isso, seguindo a tendência da economia linguística, é muito comum que essa nasalidade desapareça, como é muito frequente no PB, mesmo entre os falantes mais escolarizados:

cantaram [kã'tarãʊ] > [kã'taru]
falaram [fa'larãʊ] > [fa'laru]
fizeram [fi'zɛrãʊ] > [fi'zeru]
homem ['õmɛ̃ĩ] > ['õmi]
ontem ['õtɛ̃ĩ] > ['õŋ̃i]
bobagem [bo'baʒɛ̃ĩ] > [bo'baʒi]
passagem [pa'saʒɛ̃ĩ] > [pa'saʒi]

Cabe mencionar que em galego as palavras correspondentes a *homem* e *ontem* são *home* e *onte*, assim como *virgem* é *virxe*, e a terminação *-agem* é *-axe*.

6. ROTACISMO: consiste na substituição da lateral [l] pela vibrante [ɾ]. Essas duas consoantes são chamadas tradicionalmente de *líquidas* e têm características especiais que as aproximam dos sons vocálicos. É grande o número de línguas que só apresentam uma delas: o japonês, por exemplo, só tem [ɾ], enquanto o chinês só tem [l]. O tupi só tem [ɾ], enquanto o náuatle, importante língua indígena do México, só tem [l]. O rotacismo mais notável na história das línguas do grupo portugalego foi o que se operou nos encontros consonantais:

- *blatta* > *brata* > *barata*
- *clavu-* > *cravo*
- *gluten* > *grude*
- *plaga* > *praia*
- *planu-* > *prão* > *pordo*

- *platta* > *prata*
- *plica* > *prega*
- *plumbu-* > *prumo*
- *sclavu-* > *escravo*

Na obra de Camões (século XVI), encontramos *frauta*, *frecha*, *ingrês*, *pranta*, *pruma* etc. Nas variedades estigmatizadas do PB, o rotacismo é extremamente frequente, sendo em muitas delas o único encontro consonantal possível. A articulação de [r] depois das oclusivas (*p*, *t*, *k*, *b*, *d*, *g*) parece ser mais cômoda do que a de [l]. No italiano, os encontros consonantais latinos com [l] tiveram a consoante substituída por um [ɪ]: *fiore* ('flor'), *chiaro* ['kɪaro] ('claro'), *chiave* ['kɪave] ('chave', do latim *clavis*), *piacere* [pɪa'ʧere], ('prazer', do latim *placere*), *più* ('mais', do latim *plus*) etc. Em variedades do sardo (língua falada na ilha da Sardenha, no mar Mediterrâneo) encontramos, por exemplo, palavras como *pranta* ('planta'), *craru* ('claro'), *frore* ('flor') etc.

7. ELIMINAÇÃO DE DISTINÇÕES NÃO FUNCIONAIS: quando a língua apresenta um par de fonemas cuja distinção não apresenta uma *alta produtividade funcional*, é comum que um deles desapareça e ceda seu lugar ao mais comum e/ou de articulação mais cômoda. Isso ocorreu em francês. Nessa língua existem dois pares de vogais nasais: [ã] e [õ], [ɛ̃] e [œ̃]. As duas primeiras permitem diferenciar uma grande quantidade de palavras:

blanc [blã] ('branco')	*blond* [blõ] ('loiro')
dans [dã] ('dentro')	*don* [dõ] ('dom')
lent [lã] ('lento')	*long* [lõ] ('longo')
rang [rã] ('fila')	*rond* [rõ] ('redondo')
sang [sã] ('sangue')	*son* [sõ] ('som')

(a) Já a dupla [ɛ̃] e [œ̃] não apresenta a mesma produtividade, até mesmo porque as palavras com [œ̃] são em número muito menor dos que as com [ɛ̃] e não representam risco de confusão. As mais conhecidas são *un* [œ̃] ('um'), *lundi* [lœ̃'di] ('segunda-feira'), *commun* [ko'mœ̃] ('comum') e *parfum* [par'fœ̃] ('perfume'). Com isso, a maioria dos falantes franceses deixaram de fazer a distinção entre [ɛ̃] e [œ̃] e passaram a usar somente o [ɛ̃], que ocorre com frequência muito maior e permite distinguir um grande número de palavras:

an [ã] ('ano')	*on* [õ] ('a gente')	*un* [ɛ̃] ('um')
comment [komã] ('como')	— — —	*commun* [komɛ̃] ('comum')
faon [fã] ('cervo')	*fond* [fõ] ('fundo')	*fin* [fɛ̃] ('fim')
paon [pã] ('pavão')	*pont* [põ] ('ponte')	*pain* [pɛ̃] ('pão')
sang [sã] ('sangue')	*son* [sõ] ('som')	*sein* [sɛ̃] ('seio')
tant [tã] ('tanto')	*ton* [tõ] ('teu')	*teint* [tɛ̃] ('tez')

(b) No português arcaico, existia uma diferença de pronúncia para o que se escrevia com *j* e com *ge*, *gi*; com *x* e com *ch*; com *s* e com *ç*, *ce*, *ci*; com *-s-* e com *z*. A escrita era útil para distinguir consoantes africadas e consoantes não africadas:

azo ['adzo]	*uso* ['uzo]
encher [ẽ'ʃer]	*enxada* [ẽ'ʃada]
gente ['dʒẽte]	*jurar* [ʒu'rar]
moço ['motso]	*massa* ['masa]

A diferença entre as africadas e as não africadas, no entanto, não servia para distinguir as palavras entre si, de modo que as africadas perderam sua fricção e passaram a ser pronunciadas como as simples [s], [z], [ʃ] e [ʒ]. É curioso observar que a mesma coisa aconteceu em francês, em que as antigas fricativas [dz],[dʒ],[ts] e [ʧ] perderam sua fricção. Nas línguas ibéricas mais próximas do português, no entanto, as transformações foram em outra direção: no galego, o [ʧ] permanece, como por exemplo em *chegar* que se diz [ʧe'ɣar]. A consoante [ʒ] em galego se desvozeou, tornando-se sua homorgânica surda [ʃ], escrita com um *x*, como em *xogar*, *hoxe*, *xeral*, *xirafa*, *xeografía*, *xurar* etc. Em castelhano, a distinção entre [dʒ] e [ʒ] desapareceu, pois essas duas consoantes foram substituídas pela aspirada [x], tão característica da língua, como nas palavras *gente*, *Juan*, *general*, *hoja*, *mujer*, *girafa* etc. Também em castelhano e galego a africada [ts] sofreu mudança e se tornou a interdental [θ], escrita *ce*, *ci* ou *z*, de modo que em ambas as línguas a palavra *civilización* se pronuncia [θiβiliθa'θion]. Esse som é equivalente ao que se representa em inglês com *th*, como em *think* [θiŋk], e se pronuncia com a ponta da língua entre os dentes.

É fácil perceber, então, que a ortografia moderna do português conserva distinções que perderam sua utilidade com o passar do tempo, o que fica visível na dificuldade que muitas pessoas experimentam no momento de escrever palavras com [ʒ] e [ʃ], [s] e [z], uma vez que as grafias *gente* e *jeito*, *fechar* e *puxar*, *aço* e *isso*, *casa* e *azul* não indicam diferença de pronúncia.

(c) As fricativas [dz], [dʒ], [ts] e [ʧ] perderam sua fricção em variedades de línguas que têm essas consoantes em sua norma-padrão. É o caso, por exemplo, do andaluz, falado no sul da Espanha, em que o verbo *escuchar*, que no castelhano se pronuncia [esku'ʧar], soa [eskuʃar]. O mesmo ocorre em variedades do italiano onde *gente* ['dʒente] se pronuncia ['ʒente].

8. HAPLOLOGIA: é a eliminação de um segmento fônico quando vem repetido numa palavra. Um caso conhecido é o da eliminação de um [r] nas palavras que têm dois: no castelhano algumas formas com essa dissimilação integram a norma-padrão da língua, como *propio* e *orquesta*, o mesmo valendo para a palavra *registo* e *rasto* em português europeu. No PB, essas pronúncias são estigmatizadas, muito embora sejam frequentes também nas variedades urbanas cultas ("frustado", "prostado", "fragância", "pertubação" etc.). O latim *rostru-*nos deu *rosto*.

(c) A haplologia também pode ser *sintática*, isto é, a eliminação de segmentos sonoros idênticos ou semelhantes num grupo de palavras que se fundem numa só, principalmente na fala mais acelerada:

Nossa Senhora > [nɔsɪ̃'ɳɔra]
liberdade de ação > [liberdadʒa'sãʊ̃]
dente de leite > [dẽdʒi'leɪʧi]

9. CRASE: é a fusão de duas vogais semelhantes numa só. Na história da formação do galego (que vai dar origem em seguida ao português), ocorreu um grande número de crases. Como muitas consoantes intervocálicas latinas sofreram síncope no galego, essas supressões deram origem a muitos hiatos formados por vogais semelhantes. Num primeiro momento, essas vogais eram pronunciadas em sílabas distintas. Mais adiante, elas se fundiram numa só, facilitando a articulação pela eliminação do hiato:

colore > *coor* > *cor*
colubra > *coobra* > *cobra*
credere > *creer* > *crer*
diabolu- > *diaboo* > *diabo*
dolore- > *door* > *dor*
fine- > *fĩi* > *fim*
lana > *lãa* > *lã*
legere > *leer* > *ler*
nudu > *nuu* > *nu*
pala > *paa* > *pá*
pede- > *pee* > *pé*
populu- > *poboo* > *povo*
sedere > *seer* > *ser*
tenere > *teer* > *ter*
uidere > *veer* > *ver*

(a) Uma diferença marcante entre o PB e o português europeu é o tratamento dado às vogais resultantes de crase em cada uma das duas línguas. No português europeu, língua em que as vogais pretônicas tendem a ser pronunciadas muito fechadas ou simplesmente não são pronunciadas, as vogais resultantes de crase, mesmo na posição pretônica, são abertas:

corado [kɔ'raðu] e não [ku'raðu]
padeiro [pa'ðɐiru] e não [pɐðɐiru]
pegada [pɛ'ɣaðɐ] e não [p'ɣaðɐ]
esquecer [ʃkɛ'ser] e não [ʃk'ser]

Como a diferença entre vogal resultante de crase não é audível para os falantes de PB, língua em que não existe o [ɐ] central fechado e em que o [a] átono e o [a] tônico pouco se diferenciam, é fácil compreender a dificuldade da maioria dos brasileiros no momento de grafar o *à*. Tudo seria simples e prático se a preposição *a* fosse sempre escrita com acento — *à* —, mas as forças centrípetas que vigiam as convenções ortográficas talvez nunca venham a adotar essa solução. Com isso, o "saber usar a crase" se transforma num instrumento de seleção de alguns e de exclusão de muitos, como se um sinal de menos de meio milímetro fosse capaz de determinar a proficiência de alguém no uso da língua escrita.

10. DESASPIRAÇÃO: a letra *h* era usada pelos romanos para representar uma aspiração, por isso muitas línguas se serviram dela para o mesmo fim, como o inglês (*home, house, hand*) e o alemão (*Heim, Haus, Hand*). Quando tinham de escrever palavras gregas em que apareciam as letras θ, ρ, φ e χ, os romanos usavam uma combinação de duas letras: *th, rh, ph* e *ch*. Essa representação latina deixa evidente que, no grego antigo, aquelas letras representavam uma consoante seguida de uma aspiração, e os romanos, ao pronunciar esses empréstimos gregos, também emitiam essa aspiração. Desse modo, a palava grega φιλοσοφία era transcrita em alfabeto latino como *philosophia* e pronunciada [pʰilo'sopʰia].

(a) Para quem já estudou ou estuda inglês, basta se lembrar que, nessa língua, as consoantes *p, t, k*, quando em início de palavra, apresentam uma aspiração: a palavra *pie* ('torta') se pronuncia [pʰaɪ] e não [paɪ], como nós pronunciamos *pai*. Para um falante de inglês, a não aspiração de *p, t, k* iniciais identifica o falante como estrangeiro, por melhor que ele domine a língua.

(b) Com o tempo, no entanto, e pela pressão da economia de esforço articulatório, essa combinação do latim clássico de oclusiva com aspiração acabou se fundindo numa só consoante: *ph* passou a ser pronunciado como [f], *th* como [t] e *ch* como [k]. É por isso que hoje escrevemos *filosofia*, com *f*. A ortografia do inglês adotou o dígrafo *th* para representar a interdental [θ] em palavras como *think, thunder, moth*. Em alemão, o *ch* foi adotado para representar [χ] (semelhante ao som do *j* em espanhol: *Juan, julio, jamás*) em palavras como *machen* ('fazer'), *acht* ('oito'), *nicht* ('não').

(c) O *h* também foi adotado pelos romanos para representar a aspiração em palavras de sua própria língua, frequente sobretudo em início de palavra: *hodie, humanum, hostis, hiatus, habere* etc. Já no latim imperial, isto é, na língua usada nas diferentes províncias conquistadas por Roma, essa aspiração inicial estava abandonada. Se até hoje escrevemos *hoje, humano, hoste, hiato, haver* é para tentar manter visível a etimologia da palavra. Essa grafia etimológica com *h* também vigora em francês e em espanhol, mas não em italiano, onde se escreve *oggi, umano, oste, iato, avere*, o que decerto é muito mais lógico e menos suscetível de provocar confusão (além de economizar tinta!).

(d) A aspiração das palavras latinas desapareceu cedo no latim vulgar, portanto. Em francês, depois da conquista da Gália romana pelos francos, povo de língua germânica, o *h* voltou a representar aspiração em palavras oriundas principalmente do frâncico que, como toda língua germânica, tinha essa aspiração em início de palavra: *haine* ('ódio'), *hameau* ('lugarejo'), *hardi* ('ousado'), *haricot* ('vagem'), *hêtre* ('faia'), *homard* ('lagostim'), *houille* ('hulha'), *honte* ('vergonha') etc. Mas essa aspiração reintroduzida pelos francos foi em seguida mais uma vez descartada pelos falantes, de modo que embora até hoje se fale de um *h aspiré* ('*h* aspirado') em francês, ele na verdade é mudo. A única sobrevivência da aspiração é a não aglutinação do artigo ou do pronome à palavra: assim, se de um lado se diz e se escreve *l'homme, l'huile, l'hôtel* ('o homem', 'o óleo', 'o hotel') e *j'honore, je m'habille, j'hésite* ('eu honro', 'eu me visto', 'eu hesito'), do outro também se diz e se escreve *la honte, le hameau, le homard, je hais* ('eu odeio'), mas sem pronunciar o *h*.

(e) O castelhano se distingue das demais línguas românicas por ter eliminado o [f] no início de muitas palavras: *hoja* ('folha'), *hijo* ('filho'), *hondo* ('fundo'), *hervir* ('ferver'), *hoguera* ('fogueira'), *humo* ('fumo'), *hada* ('fada') etc. Se discute muito qual teria sido a causa da passagem de [f] para [h], mas o que nos interessa aqui é ver que o *h*, antigamente aspirado, acabou se desaspirando, como já tinha acontecido no latim vulgar com o [h] inicial do latim.

(f) Nos subúrbios de Londres se fala uma variedade de inglês chamada *cockney*, com aspectos linguísticos muito característicos e bem diferentes do inglês urbano padrão. Um desses traços é precisamente a não aspiração do *h* em palavras como *home, house, hair* etc., que os falantes de inglês padrão designam como *to drop the H's* ('deixar cair os Hs'). Essa desaspiração também ocorre em variedades do alemão e do sueco.

Os fenômenos de economia linguística não se restringem à fonologia, mas ocorrem também intensamente no plano da morfossintaxe. Aqui, o principal mecanismo cognitivo é o que opera com a eliminação das redundâncias. Esse princípio de economia linguística pode ser traduzido assim: <se determinada categoria sintática já vem expressa num elemento do sintagma, não precisa ser expressa nos demais>. O principal campo de ação desse princípio é o da concordância [▸994].

Por exemplo, no domínio da *concordância verbal*, podemos representar o princípio <evite redundância> assim:

$$\frac{\text{+morfologia verbal}}{\text{–índice pessoal-sujeito}} = \frac{\text{–morfologia verbal}}{\text{+índice pessoal-sujeito}}$$

(a) Isso significa que, se a língua tem uma morfologia verbal em que o sujeito do verbo já vem expresso pela flexão, o índice pessoal[2] de sujeito é dispensável e dispensado. Em latim era esse o caso:

amo
amas
amat
amamus
amatis
amant

(b) Por isso, nos textos latinos, é muito raro encontrar os índices pessoais do caso reto (*ego, tu, nos, vos*) junto aos verbos. Essa situação se preservou em grande medida no toscano (base do italiano oficial), no castelhano e no galego. No francês antigo (período de 850 a 1500), esse paradigma verbal com marcas bem distintas para cada pessoa também existia. Vamos ver o verbo *chanter* ('cantar') no imperfeito do indicativo daquela época:

1. chantoi*e*
2. chantoi*es*
3. chantoi*t*
4. chanti*ons*
5. chantoi*ez*
6. chantoi*ent*

(c) Para as pessoas que já estudaram francês, convém avisar que, na língua antiga, essas marcas flexionais de pessoa eram pronunciadas. Com o passar do tempo, no entanto, foi ocorrendo uma erosão fonética (fenômeno comum nos processos de gramaticalização, que vamos estudar mais abaixo) e as terminações verbais das pessoas 1, 2, 3 e 6 acabaram por se tornar idênticas. Para evitar a confusão, os falantes da língua passaram a enunciar, antes do verbo, o índice pessoal-sujeito correspondente. Essa prática se gramaticalizou, isto é, se tornou regra obrigatória na língua, de modo que no francês moderno todo verbo vem sempre precedido de seu sujeito, enunciado como um nome ou um índice pessoal. Retomando o verbo *chanter*, o imperfeito atual se conjuga assim:

1. je chantais	[ʒə ʃãtɛ]
2. tu chantais	[ty ʃãtɛ]
3. il chantait	[il ʃãtɛ]
4. nous chantions	[nu ʃãtiõ]

[2] Sobre nossa opção por chamar os "pronomes pessoais" de índice de pessoa ou índice pessoal, ver [▸463].

5. vous chantiez [vu ʃãtie]
6. ils chantaient [il ʃãtɛ]

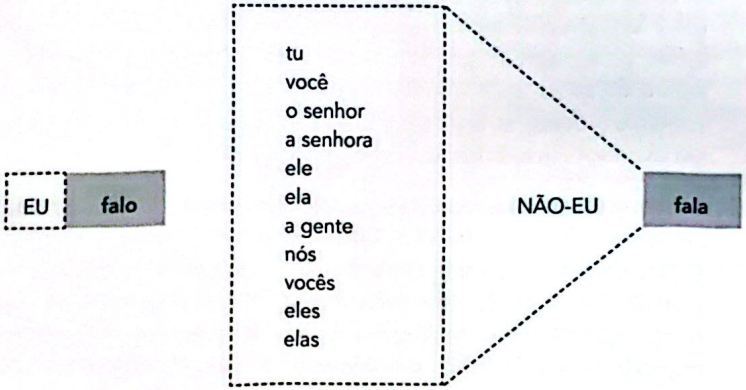
(d) Como a ortografia francesa é extremamente conservadora (refletindo o estado da língua por volta do século XVII), a forma escrita esconde o que o ouvido revela: a pronúncia idêntica do verbo conjugado nas pessoas 1, 2, 3 e 6. Observe que nas pessoas 3 e 6 não tem nada que marque na pronúncia a diferença entre o singular e o plural, de modo que somente o contexto de enunciação da fala (ou o texto escrito) pode indicar o número.

(e) O mesmo fenômeno se deu em inglês, e de maneira ainda mais radical do que em francês. No inglês antigo (falado entre 400 e 1100), havia algumas formas verbais que se distinguiam das outras pela terminação, embora já nessa época as três formas do plural fossem iguais. Ao longo da história, no entanto, as marcas flexionais de pessoa se reduziram drasticamente: no presente, somente a não-pessoa (pessoa 3) apresenta uma marca distinta, um *-s* (*she love*s), enquanto todas as outras ficam iguais ao infinitivo (*love*). No tempo passado, nem mesmo essa solitária forma flexionada sobreviveu, de modo que para todas as pessoas a forma única é *loved*. Isso exige a enunciação dos índices pessoais-sujeitos, que é obrigatória em inglês, língua em que nunca encontramos um verbo desacompanhado de seu sujeito.

(f) No português clássico, o paradigma verbal era muito semelhante ao do latim:

1. amo
2. amas
3. ama
4. amamos
5. amais
6. amam

(g) Trazido para terras brasileiras, esse paradigma passou por várias mudanças. Uma das mais importantes, segundo os pesquisadores, foi a generalização do uso do pronome *você*. Outras mudanças foram o desaparecimento total das formas correspondentes a *vós*, a introdução do pronome *a gente*, o emprego das formas *o senhor/a senhora* como marcas de tratamento respeitoso. Com isso, uma forma verbal como *falava* poderia corresponder a *eu, você, ele, ela, a gente, o senhor, a senhora*. Assim, do mesmo modo que ocorreu em francês e inglês, o PB começou a se tornar uma língua em que o sujeito do verbo tem que ser enunciado, na forma de um nome ou de um índice pessoal. Essa tendência se generalizou ainda mais, de modo que, mesmo entre os brasileiros que usam o pronome *tu*, a forma verbal mais frequente é a que corresponde à não-pessoa: *tu canta, tu cantou, tu cantava* etc., como se verifica no Rio Grande do Sul, por exemplo. Nas variedades em que ao *tu* correspondem as formas do português clássico (como no Maranhão e no Pará), existe uma acirrada concorrência entre *tu cantas* e *tu canta*. Nas variedades que tiveram sua origem histórica nas situações de *contato* entre línguas africanas e o português, o paradigma da conjugação passou por mudança ainda mais radical, já que as flexões, no presente do indicativo, se reduziram a duas: uma para *eu* e outra para *não-eu*:



(h) O linguista alemão Eberhard Gärtner, especialista em línguas românicas, observa que nas línguas bantas, as mais faladas pelos escravos trazidos para o Brasil, as marcas de pessoa e número são expressas por prefixos e não por flexões. Segundo ele, os falantes bantos podem ter identificado os índices pessoais-sujeitos do português com os seus prefixos pessoais e, ao falar a nova língua, passaram a expressar o sujeito somente pelo índice pessoal, desconsiderando e eliminando a redundância constituída pelas flexões. Vamos examinar isso em [▸242].

(i) Situação semelhante ocorreu com o africâner, língua que se formou na África do Sul com base no holandês, língua dos europeus que colonizaram o país. Nessa língua, nem sequer ocorre a distinção entre *eu* e *não-eu* que observamos no PB. Observe a diferença entre o holandês e o africâner na conjugação do verbo *zijn* ('ser'):

HOLANDÊS	AFRICÂNER
ik ben	ek is
jij / u bent	jy / u is
hij / zij / het is	hij / zij / dit is
wij zin	ons is
julle zijn	julle is
zij zijn	hulle is

(j) Um fato notável nesse caso é que nem mesmo o verbo 'ser', que em inúmeras línguas é irregular, escapou do processo de regularização do paradigma. Outro aspecto digno de nota é o emprego generalizado da forma originalmente reserva-

da para a não-pessoa (*is*) para toda a conjugação. Justamente pelo seu caráter de não-pessoa, ou seja, pelo seu caráter de não identificação com uma das pessoas do discurso, essa forma verbal tende, em muitas línguas, a se generalizar [▶464]. É o caso do PB, por exemplo, em que o passado *cantou* pode, em muitas variedades, corresponder a todas as pessoas e não-pessoas do discurso, com exceção de *eu* (*tu cantou / ele cantou / nós cantou / vocês cantou / eles cantou*). Mais uma vez, é fácil perceber que a acusação que se lança sobre os falantes das variedades menos prestigiadas de que "falam errado" ou "destroem a língua" é de ordem puramente social, política, cultural, ou seja, um mero preconceito, uma vez que o mesmo processo de simplificação do paradigma verbal se verifica em inglês, em africânder, em francês e em muitas línguas mundo afora.

(k) A eliminação da redundância no caso da concordância verbal é bastante clara: em vez de indicar a pessoa duas vezes, com o índice pessoal e com a flexão, o princípio da economia linguística se aplica, restringindo a indicação morfológica somente a um dos elementos do sintagma. É importante notar que, na comparação entre diversas línguas, a perda das marcas de flexão com o correspondente uso do índice pessoal-sujeito de maneira obrigatória parece ser preponderante. Para muitos linguistas, o processo é cíclico, já que as marcas de pessoa e número podem ser resultantes da aglutinação de pronomes muito antigos que, pelo processo de gramaticalização, perderam sua autonomia lexical e se tornaram morfemas. Por isso, o linguista americano T. Givón afirma que "a morfologia de hoje é a sintaxe de ontem".

O mesmo princípio de <evite redundância> opera na concordância nominal. Aqui também a proporção se repete:

$$\frac{\text{determinante +plural}}{\text{determinado –plural}} :: \frac{\text{determinante –plural}}{\text{determinado +plural}}$$

(a) Já mencionamos o caso do francês [▶161], em que as marcas de plural são escritas, mas nunca são pronunciadas, nem sequer nos determinantes. O PB não chegou a esse estágio mais radical (resultante, sem dúvida, das consequências do *contato* linguístico ocorrido na história da França entre gauleses, romanos e germânicos), pois o morfema *-s* de plural se preserva nos determinantes (artigos, possessivos, demonstrativos, quantificadores etc.): *as menina / minhas menina / aquelas menina / muitas menina*. Aqui também é possível ver uma consequência do contato linguístico do português com as línguas bantas como adiantamos: nessas línguas, a indicação de número é feita por prefixos e não por terminações. Os falantes bantos podem ter identificado os determinantes do português com seus prefixos de número, eliminando as marcas de plural nos nomes. [▶242]

Uma prova eloquente dessa regra está no plural do interrogativo/exclamativo em variedades mineiras e goianas, plural que desaparece do resto do sintagma: Ques *caixa é essa aqui?* Ques *menino bagunceiro!*

(b) O inglês antigo, como o latim, o grego, as línguas eslavas e as línguas germânicas atuais, apresentava um quadro de declinações, em que as palavras assumiam formas diferentes, por meio de flexões, de acordo com a função sintática que exerciam. Num processo que levou vários séculos, o sistema de declinação desapareceu por completo (em paralelo com a simplificação da conjugação verbal), de modo que nomes e determinantes perderam suas terminações de caso. O único caso sobrevivente em inglês é o genitivo, marcado pelo *'s* , como em *Peter's house* ('a casa de Pedro'). O contato do inglês com o francês (levado para a Inglaterra depois da conquista da ilha pelo nobre francês Guilherme da Normandia em 1066 e que impôs o francês normando como a primeira língua oficial do reino) fez desaparecer as marcas típicas de plural das línguas germânicas e generalizou o uso do morfema de plural *-s*, característico das línguas românicas ocidentais. Assim, o plural passou a ser marcado exclusivamente nos nomes por meio do *-s*, enquanto os determinantes permaneceram invariáveis:

sing.	plur.
the book	*the books*
my book	*my books*
some book	*some books*

(c) Poucos nomes em inglês têm seu plural formado de outro modo: *man/men* ('homem'/'homens'), *woman/women* ('mulher'/'mulheres'; mas aqui, apesar da alternância *a/e*, o plural se faz pela pronúncia do *o* como *i*); *goose/geese* ('ganso'/'gansos'), *tooth/teeth* ('dente'/'dentes'), *foot/feet* ('pé'/'pés'), *child/children* ('criança'/'crianças'), *ox/oxen* ('boi'/'bois') e, por fim, *person/people* ('pessoa'/'pessoas'), duas palavras de origem latina (*persona / populus*), ao contrário das anteriores, que são todas heranças germânicas.

(d) Por isso, os demonstrativos do inglês *this/that*, que apresentam uma forma diferente no plural, *these/those*, representam uma exceção notável, sendo os únicos determinantes da língua que se flexionam conforme o número dos determinados. Por outro lado, os adjetivos, que no inglês antigo e nas demais línguas germânicas (assim como no latim e nas línguas românicas), concordam em gênero e número com os substantivos, passaram pelo mesmo processo que os demais determinantes: não têm marcas de gênero nem de número:

my favorite thing	my favorite things
('minha coisa favorita')	('minhas coisas favoritas')

(e) A perda total da categoria de *gênero* e a redução das marcas de *número* em inglês têm seu equivalente na formação das chamadas línguas crioulas. No cabo-

-verdiano, por exemplo, o possessivo *nha* se traduz por 'meu', 'minha', 'meus', 'minhas', como em *nha companheira* ('minha companheira') e *nha sentimento* ('meu sentimento'). A atribuição do rótulo *crioula* a determinada língua tem muito mais a ver com ideologia do que com algum fenômeno propriamente linguístico. A história da Inglaterra é uma longa sucessão de invasões de povos falantes de línguas diferentes que, entrando em contato, acabaram por gerar a língua inglesa, que poderia muito bem ser classificada como uma "língua crioula" pelas semelhanças que apresenta com tantas línguas que recebem essa classificação.

(f) O contato linguístico é uma poderosa força centrífuga: ele acelera o ritmo das mudanças, que muitas vezes se tornam bem radicais. Nesses casos, é universal o processo de eliminação das redundâncias e de tudo o que não tenha uma funcionalidade óbvia, tudo o que não represente um elemento decisivo para a interação sociocomunicativa. Por isso, as línguas "crioulas" em geral não têm artigos (e as eslavas também não) e as que têm não apresentam marcas de gênero e número (como em inglês), suas palavras não se flexionam em gênero, seu paradigma verbal exige a enunciação do índice pessoal-sujeito enquanto os verbos permanecem inalterados em sua forma conjugada. Outra característica geral dessas línguas é a fixação da ordem SVO (sujeito-verbo-objeto), que também caracteriza as línguas românicas e, mais uma vez, o inglês. [▸453]

A economia linguística não deve ser confundida com a "lei do menor esforço", uma noção que acabou por se popularizar entre os não especialistas. Além dos já vistos casos de *perdas*, a economia linguística também é responsável por *ganhos*. O linguista americano Edward Sapir dizia que "toda língua tem goteiras", isto é, toda língua apresenta alguma falha ou deficiência que os falantes logo tratam de reparar.

Um bom exemplo é o caso do possessivo *seu* (e flexões). Teoricamente ele pode se referir a *você, ele, ela, vocês, eles, elas*, de modo que, também teoricamente, numa pergunta como *Você e Daniel vieram no seu carro?*, o *seu* poderia ser interpretado como "o carro que pertence a você" ou "o carro que pertence a Daniel". Como se sabe, essa ambiguidade foi resolvida com a delimitação de *seu/sua/seus/suas* àquilo que se refere a *você* (mas não a *vocês*), enquanto para a não-pessoa reservamos as contrações *dele/dela/deles/delas*. Para o plural *vocês*, no mesmo espírito de evitar ambiguidade, usamos a locução *de vocês*, de modo que perguntamos a uma pessoa: *Esse carro é seu?*, mas para duas pessoas: *Esse carro é de vocês?* Em algumas variedades do PB, como em Minas Gerais e Goiás, é comum o uso do plural *seus* para se referir a *vocês*: *Esse carro é seus?*, um uso desconhecido nas demais regiões brasileiras.

Também em inglês podemos encontrar algo semelhante. O índice pessoal *you* tem a mesma forma no singular e no plural, e como o verbo não sofre modificação de forma, poderia surgir ambiguidade na interpretação do referente desse índice. Assim, nas variedades do sul dos Estados Unidos, surgiu a forma *you-all* ou *y'all* (formada de *you* e *all*, 'todos') com seu correspondente possessivo *your-all's*. Em outras variedades encontramos outras formas como *you-uns*, *youse*, *you guys*. Com isso, a potencial ambiguidade desaparece.

É preciso ressaltar que na economia linguística não são eliminadas distinções importantes ou, quando isso ocorre, também ocorre um processo de compensação para que não haja perturbação na função sociocomunicativa da língua. Por exemplo, no PB desapareceu a distinção entre *este* e *esse*. O demonstrativo *este* sobrevive apenas na língua escrita e mesmo assim longe dos usos previstos pela tradição normativa [▸792]. O desaparecimento de *este* se deve a um processo de assimilação [▸793]. Para compensar essa perda, os falantes passaram a empregar *esse* acompanhado de advérbios de lugar: *esse aqui/esse aí*, recuperando a propriedade dêitica do demonstrativo. Essa combinação de demonstrativo com advérbio de lugar também se verifica em muitas outras línguas (francês, catalão, alemão, para citar só essas).

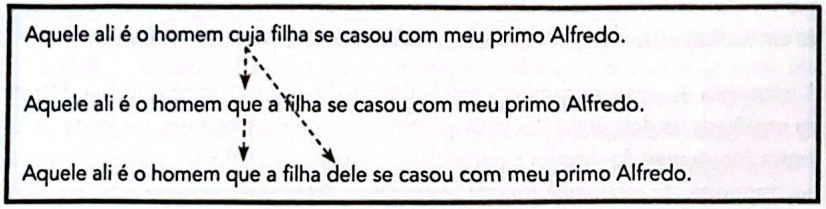
A economia de recursos também incide sobre o domínio dos *tempos verbais*. Diante da existência de dois *pretéritos mais-que-perfeitos*, um simples e um composto, os falantes abrem mão do simples e empregam, na interação oral, exclusivamente o tempo composto, de modo que formas como *fizera, faláramos, conhecêramos, perdoara* etc. nunca ocorrem na fala espontânea, a não ser com objetivo humorístico. Além disso, diante da possibilidade de uso de dois verbos auxiliares para a formação do mais-que-perfeito composto, os falantes dão preferência a *ter*, de uso mais amplo do que *haver*, reservando *haver* para gêneros escritos mais monitorados ou para eventos de fala formais ou "hipercorretos".

O mesmo ocorre com o chamado *futuro do pretérito* que, graças à coincidência das terminações, é amplamente substituído nas interações orais pelo *imperfeito do indicativo*, sobretudo na expressão das hipóteses: "Se eu tivesse condições, *comprava* um carro novo", em lugar de *compraria*. A forma *gostaria* sobrevive como item cristalizado, estereotipado. Como o uso da mesma forma verbal se distribui por contextos semântico-pragmáticos bem delimitados, não há risco de ambiguidade na interpretação dos enunciados: "Você *podia* me ajudar a limpar a sala?" / "Naquela época você não *podia* imaginar que ele era tão mesquinho". Cabe também observar que a expressão do *futuro do pretérito* se faz com muita frequência por meio de uma forma composta com o emprego do auxiliar *ir*: "Se você experimentasse, tenho certeza que *ia gostar* de bacalhau". Fenômeno semelhante ocorre em francês, em que o *imperfeito do subjuntivo* jamais é empregado na língua falada e rarissimamente na escrita, tendo sido substituído amplamente pelo *imperfeito do indicativo* em todas as circunstâncias.

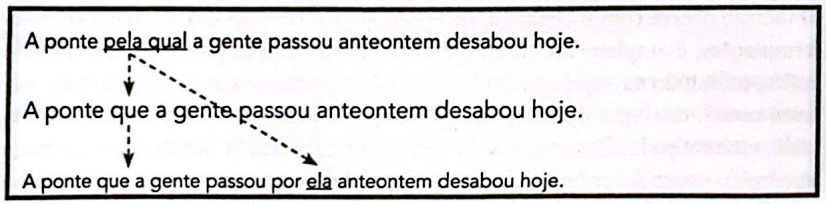
11. ANÁLISE VERSUS SÍNTESE [▸445]: as línguas românicas se formaram com base no latim vulgar, no qual já imperava a substituição das formas *sintéticas* pelas formas *analíticas*, o que é um claro procedimento de economia linguística, pois desobriga a memória de conservar uma grande quantidade de formas flexionadas. Esse processo

continua vivo e ativo nessas línguas, principalmente em suas variedades menos sujeitas às forças centrípetas. A eliminação total do sistema de *declinação*, substituído pelo enrijecimento da ordem dos constituintes na sentença e pelo uso intenso de preposições, é a evidência mais contundente desse processo. O abandono das formas verbais sintéticas, caracterizadas pela flexão — *falavam* — em favor das analíticas — *eles falava* — também integra esse processo, tendo se tornado, em francês, a única opção disponível: *je / tu / il / elle / on / ils / elles* [parlɛ], o mesmo valendo para o inglês, língua também essencialmente analítica.

O desaparecimento de *cujo* na interação oral em PB, substituído pelo *que*, chamado de *relativo universal*, sozinho ou em combinação com expressões possessivas, é um nítido exemplo da vitória da análise contra a síntese:



Esse processo incidiu, de fato, sobre todas as construções com conectivo relativo [▶899]. As orações que o verbo delas rege sintagmas preposicionais deixaram de apresentar a preposição antes do relativo: o lugar da preposição pode ficar vago ou então a preposição é deslocada para depois do verbo, seu lugar natural na língua, seguida de um pronome-lembrete:



Esses processos de mudança nas construções com relativo atingem outras línguas da família românica, não sendo portanto exclusividade do PB. O linguista francês Pierre Guiraud (1973: 48) oferece os seguintes exemplos de sua língua:

Tu m'as envoyé le colis que tu me parlais.
('*Você me enviou o pacote que você me falava*')

Mon mari que je suis toujours sans nouvelles.
('*Meu marido que eu ainda continuo sem notícias*')

Mon mari que je suis sans nouvelles de lui.
('*Meu marido que eu estou sem notícias dele*')

Une chose que tu peux être fier.
('*Uma coisa que você pode ficar orgulhoso*')

No galego moderno, em registro menos monitorado, também ocorre o fenômeno de apagamento da preposição ou de devolução da preposição a seu lugar depois do verbo, com pronome-lembrete (exemplos dados por Álvarez e Xove, 2002: 578):

Alicia é unha rapaza que *estudiara Pancho* con ela.
('Alicia é uma moça que Pancho tinha estudado com ela')

E axiña chegaron a un sitio que *había unha cruz* nel.
('E depressa chegaram num lugar que tinha uma cruz nele')

Os mesmos autores assim se referem ao relativo *cuxo* ('cujo'):

É um relativo antigo, desaparecido da língua espontânea moderna e não contemplado pela normativa padrão. [...] A língua moderna eliminou o relativo *cuxo* e a sua construção. (2002: 590).

No galego, portanto, as construções onde deveria ocorrer *cuxo* são realizadas exatamente como no PB, pelo desmembramento do relativo em *que* seguido ou não de uma expressão possessiva. O desaparecimento de *cuxo* em galego se torna ainda mais patente quando verificamos que no *Diccionario da Real Academia Galega* a palavra *cuxo* não está registrada como relativo.

No castelhano, o fenômeno se repete. O linguista espanhol Emilio Lorenzo (1994: 232) cita os seguintes exemplos, colhidos por ele no espanhol falado e escrito em Madri, que também demonstram o desaparecimento do *cuyo*:

El padre que su *hijo vino ayer.*
('O pai que o filho veio ontem')

En una casa que su *amo es médico.*
('Numa casa que seu dono é médico'].

El gran cantante que su *llegada las admiradoras de España tanto han esperado.*
('O grande cantor *que sua* chegada as admiradoras da Espanha tanto esperaram')

Apesar de seu desaparecimento na língua falada, considero que ainda cabe ensinar o uso de *cujo* para a produção de GTM [▶906]. Mesmo assim, é preciso ter consciência de que, mesmo na escrita monitorada, o pronome tende a desaparecer:

(1) Foi morar fora, mas escolheu um país que a bandeira é vermelha e branca (Canadá!). [colunas.globoesporte.com/miltonneto/2008/12/01/que-vitoria/, acesso em 16/1/2009]
(2) Cito, dentre outras, 3 comissões de inestimável valor para a ordem atual e o futuro do clube. A comissão de orçamento, que o nome já diz tudo [...] [oglobo.globo.com/blogs/flamengo/posts/2009/05/06/pilulas-rubronegras-x-183635.asp]
(3) Não somos um país autônomo, não somos um país que tenha um mercado de trabalho interno fortalecido, não somos um país que o governo tenha uma autonomia econômico-financeira, não somos um país que tenha condições de suportar um quadro recessivo internacional, se ele vier numa dimensão mais profunda. [www.corecon-rj.org.br/entrev_det.asp?Id_ent=33]
(4) Muitos motoristas já ouviram falar, já viram, ou foram vítimas mesmo, de um carro que o motor ferveu. [www.tudonahora.com.br/noticia.php?noticia=38669]
(5) Era uma vez uma menina que o pai dela casou pela segunda vez. Ela perdeu a mãe muito cedo, aí o pai casou, deu madrasta. (Altimar de Alencar Pimentel, *Contos populares de Brasília*. Brasília: Thesaurus, s/d, p. 28)

(6) Outro dia, alguém que acabei não anotando o nome sugeriu que as Nações Unidas fossem dotadas de uma força de intervenção rápida permanente, capaz de atuar em casos agudos (como o de Honduras) ou que já se tornaram crônicos (Darfur, por exemplo). (Clóvis Rossi, *Folha de S.Paulo*, 26/7/2009)

4.4.2 A *gramaticalização*

Os estudos acerca dos processos de gramaticalização têm se desenvolvido muito nas últimas décadas. Cada vez mais, os linguistas percebem as complexidades e as consequências profundas da gramaticalização na vida das línguas e de seus falantes. Aqui, vamos nos concentrar em alguns processos da gramaticalização que explicam mudanças ocorridas e ocorrendo nas línguas humanas. E vamos definir *gramaticalização*, de modo muito grosseiro, como a produção de novos recursos gramaticais a partir de (re)processamentos cognitivos, por parte dos falantes, impostos aos recursos gramaticais já existentes. Desses processamentos cognitivos o mais destacado é a metáfora.

A língua não é uma entidade autônoma, que pode ser estudada sem que se leve em conta os falantes e suas interações sociocomunicativas. Sem falantes não existe língua, e a língua só existe *em uso*, na forma de *discurso*. Repito aqui a imagem que empreguei na introdução, inspirada no símbolo taoísta do *yin* e do *yang*:

Assim, a gramática da língua se forma a partir dos usos que os falantes fazem dos recursos verbais que estão à sua disposição no sistema. Se esses recursos se revelam insuficientes, os falantes, em sociedade, criam novos recursos, quase sempre através de reestruturações, ressignificações, reinterpretações e reanálises dos recursos já existentes. Gramática e discurso estão em íntima conexão, e um já traz dentro de si os germes ou as sementes do outro.

Uma vez que a função primordial das línguas humanas é permitir a interação de seus falantes por meio da produção de sentidos, os recursos linguísticos se vinculam todos ao campo da semântica, que é precisamente o campo da significação. Falar é dar sentido ao mundo real, ao mundo virtual e ao mundo do discurso; é simbolizar a experiência vivida ou imaginada; é organizar a interação sociocomunicativa com

vistas ao compartilhamento dos significados construídos conjuntamente, graças a essa mesma interação.

Mas os recursos linguísticos, sozinhos, ficam confinados a um universo de potencialidades abstratas, de significações virtuais. Parafraseando um conhecido poema de Carlos Drummond de Andrade ("Procura da poesia"), poderíamos dizer que, no domínio da semântica, os recursos linguísticos se encontram "em estado de dicionário", latentes, imóveis, à espera: cada um deles "tem mil faces secretas sob a face neutra". Quando ativados pelo falante, no entanto, quando trazidos para dentro do discurso — que é a única manifestação verdadeira da língua —, essas faces secretas assumem um rosto definido, dão a cara a tapa, essa neutralidade se dissipa e dá lugar a uma força direcionada, a uma tomada de partido, que permite a produção de sentido em contextos reais de interação. Gramática e discurso, portanto, se conectam e se interconectam contra o pano de fundo da semântica, ativada pelos motores da pragmática.

O mais importante processo cognitivo envolvido na gramaticalização, como afirmei acima, é a *metáfora*. O próprio uso que fazemos da palavra *metáfora* é metafórico. Em grego, *metaphorá* significa 'transporte para outro lugar', no sentido mais concreto, tanto que os caminhões de mudança na Grécia moderna trazem a palavra ΜΕΤΑΦΟΡΕΣ (METAFORES) estampada em grandes letras coloridas. Os romanos traduziram o verbo grego *metaphorein* por *transferire*, com o mesmo significado, inicialmente concreto, de 'levar para outro lugar'.

A metáfora é, portanto, o processo mental que fazemos ao *transferir* uma palavra de um domínio semântico para outro, ao *expandir* seu significado de um sentido concreto para sentidos cada vez mais abstratos. Na gramaticalização, esse processo se dá primordialmente pela transferência de elementos do léxico para a gramática, da coisa para a não-coisa, do universo empírico para o universo do discurso, do sensível para o cognoscível. O linguista Guy Deutscher, especialista em linguística histórica, ao tratar da metáfora, escreveu:

> E a pessoa, sem dúvida, não precisa ser um aspirante a poeta para transferir conceitos de uma domínio linguístico para outro. Até mesmo no discurso mais lugar-comum, é quase impossível arriscar alguns passos sem tropeçar em dúzias de metáforas. Pois as metáforas estão por toda parte, não só na língua, mas também em nossa mente. Longe de ser uma rara centelha do gênio poético, o dom maravilhoso de uns poucos privilegiados, a metáfora é um elemento indispensável nos processos mentais de cada um de nós. [...] utilizamos metáforas não por causa de qualquer pendor literário ou ambição artística, mas muito simplesmente porque a metáfora é o principal mecanismo através do qual podemos descrever e mesmo apreender a abstração. (Deutscher, 2005: 117)

As metáforas estão presentes em qualquer trecho de língua falada ou escrita, mesmo a mais corriqueira, mesmo o que trate do assunto mais árido e menos poético. Abrindo um jornal ou revista ao acaso, as metáforas pululam diante de nós. Ou melhor, diante do especialista, porque as metáforas — justamente pelo processo de gramaticalização — se cristalizam de tal maneira que se tornam elementos banais, que não despertam a atenção de ninguém.

É possível dizer que a metáfora está presente no discurso em três níveis de profundidade:

na superfície do texto, por meio de comparações, usos figurados, jogos de palavras etc.

no léxico (os usos atuais das palavras são quase sempre derivações metafóricas de seus étimos)

na gramática

Um gênero textual que parece estar livre da acusação de ser "poético" é, sem dúvida, a manchete econômica. No entanto, as metáforas estão todas ali, cristalizadas, mesmo em poucas linhas de texto:

> A CRISE[1] ECONÔMICA[2] MUNDIAL[3] ESTÁ[4] LONGE[5] DE ACABAR[6]. E AS MEDIDAS[7] DE SOCORRO[8] TOMADAS[9] PELOS GOVERNOS[10] DOS PAÍSES[11] RICOS[12] TRAZEM[13] O RISCO[14] DE PROVOCAR[15] UMA INFLAÇÃO[16] GLOBAL[17] A PARTIR[18] DE 2015. (terra.com.br, 25/7/2009)

(i) *metáforas na superfície do texto:*

- '*está longe de acabar*' → *acabar* não é um lugar, portanto nada poderia, concretamente, estar 'longe de acabar';
- '*medidas ... trazem o risco*'→ *risco* não é um objeto palpável, concreto, que possa ser 'trazido'; *medidas* não tem traço [+animado], portanto não poderia ser sujeito do verbo *trazer*, que implica 'carregar' fisicamente;
- '*provocar uma inflação*' → *provocar* é 'chamar para fora' e *inflação* é 'inchaço', logo não se poderia 'chamar um inchaço para fora', a não ser, é claro, metaforicamente.

(ii) *metáforas no léxico:*

1. crise: do grego *krisis*, termo usado na medicina antiga para indicar o momento decisivo para o tratamento de uma doença; daí o sentido metafórico de 'momento decisivo, difícil';
2. economia: do grego *oikos*, 'casa', e *nomia*, 'administração';
3. mundo: do latim *mundu-*, 'cosmético, produtos para a beleza da mulher'; há também em latim o adjetivo *mundus*, 'limpo, asseado' (de onde vem o nosso *imundo*, 'sujo'), metaforicamente: 'organizado';
4. estar: do latim *stare*, 'ficar de pé';
5. longe: do latim *longe*, 'ao longo de, em comprimento';
6. acabar: de *a* + *cabo* + *ar*; 'cabo', por sua vez, provém do latim *caput*, 'cabeça'; 'acabar', portanto, significa, concretamente, 'cortar o cabo ou a cabeça de algo';
7. medida: 'avaliação de grandeza física; medição de volume ou quantidade'; metaforicamente, 'decisão tomada após avaliação (medição) de um problema';
8. socorro: derivado de 'socorrer', do latim *sub* ('sob') + *correre* ('correr'); portanto, 'correr por baixo' e, daí, 'vir em auxílio, amparar';
9. tomar: 'tirar (algo) de (alguém) e apossar-se desse algo; subtrair, arrebatar, usurpar';
10. governo: do latim *gubernu-*, e esse do grego *kybern-* 'dirigir um barco';
11. país: do latim *pagus*, 'marco fixado na terra';
12. rico: do germânico *reiks*, 'poderoso' (correlato do latim *rex*, 'rei');
13. trazer: do latim *trahere*, 'arrastar, puxar com dificuldade';
14. provocar: do latim *pro-* + *voco* (de *vox*, 'voz'); 'chamar para fora, intimar a sair';
15. risco: talvez do latim medieval *risicu-*, de *reseco*, 'cortar, suprimir', de *re-* + *seco*, 'cortar';
16. inflação: do verbo *inflar*, 'inchar', do latim *in-* + *flo*, 'sopro';
17. globo: do latim *globu-*, 'bola, esfera', também 'formação de tropas muito apinhadas', 'multidão, chusma'; *globus lunae*: 'o disco da lua'; *globus sanguinis*, 'coágulo de sangue', *globus flammorum*, 'turbilhão de chamas';
18. partir: do latim *partire*, 'cortar, dividir, separar em pedaços'.

(iii) *metáforas na gramática*:

- *a, o* → os artigos definidos provêm de demonstrativos latinos: *illu-*, 'aquele'; *illa*, 'aquela'; reduzidos a artigos, tiveram sua carga semântica enfraquecida e sofreram erosão fonética;

- *uma* → os chamados 'artigos indefinidos' provêm dos numerais latinos *unu-*, *una*, que perderam sua significação de quantificador específico (uma unidade) para se transformar em indicadores de quantidade vaga, imprecisa (a ponto de serem usados no plural *uns*, *umas*, o que não é permitido aos numerais propriamente ditos).

Esse processo cognitivo de metaforização segue sempre um mesmo percurso que podemos resumir como do real para o virtual, do mundo que podemos apreender com nossos sentidos para o mundo apenas inteligível. Esse percurso se manifesta sob diversas formas:

1. Do concreto para o abstrato: palavras que originalmente tinham um sentido exclusivamente concreto assumiram sentidos abstratos ao longo do tempo, a ponto de só serem usadas assim na atualidade:

(a) CASO → em latim, *casu-* é o particípio passado do verbo *cadere*, 'cair', portanto equivale a 'caído'. De algo concreto que realmente cai por força da gravidade, passou a significar algo que ocorre, que acontece, algo que "cai" nas nossas vidas. O termo foi usado, já em latim, para designar as diferentes formas que uma palavra podia assumir segundo a função sintática que exercia na frase, porque, nessa visão metafórica, a palavra "caía" de sua função mais importante, a de sujeito, para funções menos importantes. Por isso, se falava em "declinação", ou seja, em queda progressiva. A extensão de sentidos foi ainda mais longe: hoje se pode dizer que alguém "está de caso" com outra pessoa, sem que a ideia de "cair" nem de longe esteja no pensamento de quem fala.

(b) CHEGAR → do latim *plicare*, 'dobrar', de *plica*, 'prega, dobra'. Supõe-se que o verbo remeta à prática de se dobrar as velas dos navios ao se chegar no porto. O italiano *arrivare* e o francês *arriver*, que traduzimos por 'chegar', derivam de *ad* + *ripa*, 'margem de rio', e também revelam uma metáfora náutica: chegar à margem, atracar na costa. É o mesmo sentido de *abordar*, 'chegar à borda'.

(c) DELIRAR → do latim *delirare*, 'escapar do sulco aberto pelo arado' e daí para 'desviar-se do caminho' e 'tresvariar'.

(d) ESTAR → em latim, *stare* era 'estar de pé'. Daí, por extensão de sentido, algo que pode se mover, que não é permanente. Da diferença entre 'estar sentado' e 'estar de pé' é que provém a distinção que fazemos entre *ser* e *estar* quando dizemos que uma pessoa *é* doente ou que *está* doente.

(e) ESTILO → do latim *stilu-*, 'utensílio formado por uma haste pontiaguda' usado para escrever sobre tábuas de cera ou argila. Daí passou a significar o próprio modo de escrever, o estilo. Do mundo da escrita, o estilo migrou para o mundo da moda e das demais artes: estilo musical, estilo de pintura etc. Em francês stylo significa 'caneta'. *Caneta*, por sua vez, é diminutivo de 'cana'.

(f) MAS → nossa conjunção *mas* tem a mesma etimologia do advérbio *mais*, ou seja, o latim *magis*. Quando digo: *Sílvia é bonita, mas é antipática*, a ideia original era dar a entender que Sílvia é *mais* antipática do que bonita. Pelo processo de gramaticalização, *mais* evoluiu para esse emprego adversativo e perdeu seu *i*, tornando-se *mas*. Observe-se que em francês a conjunção se escreve *mais* [mɛ], em castelhano a forma *mas* é a conjunção, enquanto o advérbio é *más*; em galego se usa *mais* para as duas coisas. E na língua escrita mais antiga também se usava a forma escrita *mais*.

(g) NEGÓCIO → do latim *nec* ('negação') e *otium*, 'ócio'. O negócio é o oposto do ócio, é a ocupação, o trabalho. A mesma ideia está presente no inglês *business*, derivado de *busy*, 'ocupado'.

(h) PENSAR → em latim, *pensare* era 'avaliar o peso' de alguma coisa. Para o nosso sentido de 'pensar', se dizia *cogitare* em latim. Hoje só empregamos o verbo no sentido abstrato de 'submeter algo ao processo de raciocínio lógico'.

(i) QUERER → em latim, *quaerere* era 'buscar uma coisa, procurá-la'. Para o nosso sentido de 'querer', se usava em latim o verbo *uelle*. Do particípio passado de *quaerere*, *quaestu-*, deriva a palavra *quaestione-*, 'busca, procura', de onde se originou *questão*.

(j) RIVAL → do latim *rivale-*, 'ribeirinho, que tira água do mesmo ribeiro', portanto, que disputa a posse da água, que compete por ela.

(k) SABER → em latim, *sapere* significava 'ter gosto de, ter sabor de'. Para a nossa ideia de 'saber', o latim usava *scire* (de onde vem *scientia*, 'ciência'). Observe que a palavra *sabor* guarda semelhança com o verbo *saber*, e não é por acaso.

(l) SER → em latim, o verbo *sedere*, origem do nosso *ser*, significava 'estar sentado'. Daí provém *sede*, 'o assento', que também está na origem de *sé*, 'a cadeira do bispo', *sedentário*, 'aquele que está sempre sentado' etc. O verbo 'ser' em latim era *esse*.

(m) TER → em latim, o verbo *tenere*, origem do nosso *ter*, significava 'segurar com a mão'. Para a ideia de posse, os falantes de latim usavam *habere*, origem do nosso *haver*, que também, como se vê, passou a ser usado abstratamente.

(n) TRABALHO → do latim *tripaliu-*, de *tri*, 'três', e *paliu-*, 'pau'. Nome de um instrumento de tortura composto de três estacas. Dispensa comentários...

2. Do espaço para o tempo: palavras que, inicialmente, indicavam *situação física concreta* passaram a designar *situação imaginária, abstrata, tempo*, e daí para *situação dentro do discurso*:

(a) LOGO → do latim *loco*, 'lugar'. Hoje em dia só empregamos essa palavra com sentido temporal: momento imediato, sem demora (*Faça logo o que estou mandando*); algum momento não muito distante do da enunciação (*Te encontro logo mais no cinema*). A abstratização foi mais longe e também empregamos *logo* com o sentido de 'justamente', 'ainda por cima': *Você foi emprestar seu carro logo para o Vicente, aquele irresponsável?*; e também com o sentido de 'portanto', 'em consequência': *Penso, logo existo*.

(b) ONDE → do latin *unde*, 'de onde'. Para a indicação de 'lugar em que', se usava em latim *ubi*. O uso de *onde* para indicar *tempo*, embora condenado pelos prescritivistas, está fartamente documentado na escrita mais monitorada como nos exemplos abaixo. Observe-se que em francês o pronome relativo *où* ('onde') é o único que se pode usar com referência a tempo.

(7) Ainda vivemos em um tempo onde os pacientes que nos procuram se entusiasmam e valorizam os equipamentos que ainda não viram em outros consultórios. (Paulo Schor, professor adjunto do Departamento de Oftalmologia da Universidade Federal de São Paulo. *Revista Brasileira de Oftalmologia*, 2010: 69 (1), 5-6).

(8) Para se ter uma ideia das relações entre o popular e certos hábitos cultos, basta dizer que ler 'em voz baixa' era considerado estranho num tempo onde a vida social e coletiva era predominante. Ler, segundo Zumthor, significava ler em voz alta. A leitura silenciosa era inclusive mal vista: "Homens barbudos que falavam sozinhos segurando nas mãos folhas de tecido branco". (Ricardo Azevedo, escritor, ilustrador, pesquisador. *Como o ar não tem cor se é azul? Vestígios dos contos populares na literatura infantil*. Dissertação, FFLCH, Universidade de São Paulo, 1997, p. 85).

(9) É só no atravessamento pela castração que o sujeito fará a passagem do ser para ter o falo. Algo do objeto é cedido aí. Implica numa perda de gozo. Ele passa à construção do fantasma, o que difere de um tempo onde se pensava, se oferecia como objeto falo imaginário. Neste tempo da construção do fantasma trata-se é de a-prender. [...] A entrada da criança na leitura e na escrita formal, isto é, na alfabetização, acontece num tempo onde há um fracasso das teorias sexuais, já que não consegue respondê-las. [...] A latência é o tempo onde se abre para o sujeito o aprender a aprender, a processar o que tomou do campo do Outro. (Leila Neme, "O a-prender na constituição do sujeito". *www.escolaletrafreudiana.com.br/UserFiles/110/File/.../008.pdf - acesso: 6/9/2010)*

(c) ESPAÇO → essa palavra também é empregada para indicar um período de tempo: No *espaço de um mês, todos os destroços foram recolhidos*.

(d) IR → o verbo *ir* tem o significado original de 'deslocamento no espaço': *vou para São Paulo amanhã; Ana vai para casa às oito*. No entanto, como todo deslocamento espacial implica, inevitavelmente, numa passagem de tempo, o verbo *ir* passou a ser utilizado (em quase todas as línguas do mundo) como um deslocamento também no tempo, e não mais só no espaço — o que prova que os falantes, intuitivamente, sempre souberam que tempo e espaço são uma coisa só, como viria a defender, no século XX, o físico Einstein com sua teoria da relatividade. Por isso, o verbo *ir* se tornou o verdadeiro elemento de formação do tempo futuro em nossa língua: *não vou jogar bola amanhã*. É esse uso metafórico de ir que nos permite dizer: *hoje vou ficar em casa descansando* — um aparente paradoxo, se fôssemos levar em conta o significado original de *ir*: ou se vai ou se fica. Da mesma forma, uma pessoa que já está deitada na cama, pode simplesmente se virar para o lado e dizer: *vou dormir*, sem que isso envolva nenhum deslocamento espacial, nenhuma ida propriamente dita. Também em *vou voltar daqui a pouco*, os verbos *ir* e *voltar* não se opõem. Por isso não se justifica a proibição de uso de formas como *vou* ir ou *vou vir*, porque nesses usos a forma *vou* é um mero auxiliar, uma partícula que indica o futuro das ações.

(e) A ideia de 'ir', o deslocamento espacial, é uma das mais importantes para nossa apreensão e expressão da experiência de mundo. Por isso, o verbo que a representa se transformou numa partícula gramatical amplamente empregada por inúmeras línguas do mundo com os mais diversos valores. É, sem dúvida, um dos melhores exemplos de gramaticalização. Observe os seguintes empregos (extraídos de Heine e Kuteva, 2002):

(i) *andativo* ou *itivo* (deslocamento para longe do falante):
◗ haitiano:
Voye msyea ale.
mandar homem ir
"Mande o homem embora".

(ii) *mudança de estado*:
◗ inglês:
She went mad.
ela ir louco
"Ela ficou louca".

◗ haitiano:
Madâm -la va- rich.
Senhora a ir rico.
"A senhora vai ficar rica".

(iii) *contínuo*:
◗ português brasileiro:
Vou andando por aí.

◗ wichita (Estados Unidos):
wit- i:ya:
ferver ir
"Está fervendo".

(iv) *hortativo* (imperativo):
◗ português brasileiro:
Vamos acabar com isso!

◗ baka (Sudão):
gó ja ndɔ́!
ir pegar banana
"Pegue bananas!"

(v) *alativo* (indicador de direção):
◗ fá d'ambô (ilha de Ano Bom):
wan namín zugá wan budu ba zinál
art. menino jogar art. pedra ir janela
"O menino jogou uma pedra na janela"

(vi) *futuro*:
◗ bari (Sudão):
nan tu kɔn
eu ir fazer
"Vou fazer".

- soto (Lesoto):
kē- ěa-rèka
eu ir comprar
"Estou quase comprando" / "Estou indo comprar"
- zulu (África do Sul):
ba- ya- ku- fika.
eles ir INFINITIVO chegar
"Eles vão chegar".
- margi (Nigéria):
nì àrá wì
eu ir correr
"Vou correr"
- igbo (Nigéria):
ó gà àbyá
nós ir vir
"Nós vamos vir"
- francês:
Je vais aller au cinéma
eu ir ir ao cinema
"Eu vou ir ao cinema" / "Eu irei ao cinema"

(f) Em todas as línguas conhecidas, as palavras gramaticais (advérbios, conjunções, preposições) empregadas com sentido *espacial* também são empregadas com sentido *temporal*:

ESPAÇO	TEMPO
de Belém a São Luís	*de* segunda a domingo
no Brasil	*no* Natal
à beira do abismo	à tardinha
antes do rio	antes do anoitecer
atrás do morro	mil anos *atrás*
perto da casa	*perto* da meia-noite
país *distante*	época *distante*
longe da cidade	já vai *longe* aquele dia
depois da ponte	*depois* do feriado
dentro da igreja	*dentro* de dois anos
desde a portaria do prédio	*desde* ontem
em torno da praça	*em torno* das dez horas
ali na calçada	*ali* pelas três horas
cerca de mil metros	*cerca de* mil dias
logo na entrada da cidade	*logo* depois de nascer
por aqui	*por* ora

Essas palavras gramaticais originalmente indicavam espaço e logo, por efeito metafórico, passaram a ser usadas também para designar o tempo.

(g) Observe que palavras que já surgiram como indicadores de tempo — *agora, hoje, amanhã, quando, tempo, hora* — não são usadas com sentido espacial, pois a metáfora atua sempre numa só direção: do concreto/espaço para o abstrato/tempo. Por isso é que *tempo* não designa espaço, mas *espaço* pode designar tempo; a palavra *período*, que usamos exclusivamente com designação temporal, provém do grego *peri* + *hodos*, 'o caminho em volta, circuito', assim como *época*, do grego *epokhe*, 'lugar de parada, interrupção'; a palavra *intervalo* vem do latim *intervallu-*, 'espaço entre duas estacas de uma fortificação'.

(h) Do adjetivo *duro*, usado para designar coisas concretas, temos os verbos *durar* e *perdurar*, a preposição *durante*, o substantivo *duração*, o adjetivo *duradouro*, todos usados apenas com referência ao tempo.

(i) O latim *statione-* tinha vários significados: 'estado de repouso; morada, paragem; posto, ancoradouro; lugar de reunião; estrebaria, cavalariça; lugar, sítio'. Hoje também usamos *estação* para designar as quatro partes em que o ano é dividido segundo fenômenos atmosféricos característicos.

(j) Os adjetivos *longo, largo, comprido, curto, alto* e *baixo*, medidas de dimensão espacial, também se aplicam a referências temporais — *um longo inverno, um largo momento, um verão curto, altas horas, baixa temporada* —, mas *lento, rápido, vagaroso, apressado, veloz* e outros adjetivos que indicam duração e velocidade nunca se aplicam ao espaço.

(k) Em galego, a palavra *daquela* se tornou um advérbio com o sentido de 'naquele tempo'.

(l) Da preposição *atrás*, que indica posição no espaço, formamos os verbos *atrasar* e *retrasar*, usados somente com referência ao tempo.

(m) A palavra *lapso* provém de *lapsu-*, particípio passado do verbo latino *labor*, 'escorregar, desequilibrar-se'. Daí passou ao sentido de 'erro, falha' e, em seguida, a 'intervalo de tempo'.

3. DO MUNDO EMPÍRICO PARA O MUNDO DO DISCURSO: palavras que inicialmente se referem a elementos do mundo real passam a servir como organizadores do fluxo discursivo. Esse fenômeno recebe o nome de discursivização:

(a) AÍ → nas narrativas orais, o advérbio *aí*, usado originalmente para indicar o espaço físico próximo da pessoa com quem se fala, é empregado para expressar o desdobramento dos fatos narrados: "Ontem eu encontrei o João na barbearia e a gente ficou conversando. *Aí*, chegou o irmão dele e começou a brigar comigo. *Aí*, o João tentou me defender e levou um soco do irmão. *Aí*, o barbeiro ameaçou chamar a polícia".

(b) ONDE → além de sua passagem de lugar para tempo, *onde* é cada vez mais empregado, no PB, como organizador do discurso, referindo-se a uma situação/um fato/um lugar/um tempo/um enunciado expresso anteriormente: "Aquele projeto levou cinco anos e deu um trabalhão pra gente, *onde* o grupo sabia desde o início que as verbas nunca iam ser suficientes".

(c) NISSO → contração demonstrativa que perde seu caráter dêitico empírico (mostrar alguma coisa no mundo real) e passa a indicar tempo e lugar na narrativa:

"Nós todos estávamos vendo televisão ontem à noite. *Nisso*, o telefone tocou e eu fui atender".

4. DO LÉXICO PARA A GRAMÁTICA: é o caso clássico de gramaticalização, considerado por muitos linguistas como a gramaticalização por excelência, chamando a atenção, porém, para a tendência atual de muitos linguistas de não separar rigidamente o léxico da gramática, preferindo falar de lexicogramática [▸436]. A interação entre esses dois componentes da língua é íntima e complexa, não admitindo uma divisão nítida (aliás, se essa divisão existisse, o próprio fenômeno da gramaticalização não ocorreria). Nesse processo, uma palavra autônoma se transforma num elemento gramatical, a ponto de se tornar um morfema preso e se transformar em regra obrigatória. É o que T. Givón quer dizer ao afirmar que "a morfologia de hoje é a sintaxe de ontem", ou seja, palavras que tinham liberdade e autonomia na estruturação sintática, além de precisão semântica, se transformam em elementos gramaticais, com lugares predeterminados para ocupar e com semântica fraca, atenuada. Quando uma palavra autônoma se transforma em item gramatical, além de ter sua semântica atenuada e de se fixar em lugares precisos na morfossintaxe, ela também sofre *erosão fonética*, isto é, perde material sonoro devido a seu uso intenso, como nos casos abaixo:

- cadê → esse pronome interrogativo tem sua origem em *que é de?*, com o sentido de *que é feito de x?* Ao se gramaticalizar, ele assumiu outras formas, além de *cadê*: *quedê* e *quede*. Processo semelhante ocorreu em latim: a locução *quo modo* ('de que modo') se aglutinou em *quomodo* e está na origem das formas românicas *come* (italiano), *comme* (francês), *como* (galego, castelhano, português).
- cê → o pronome *cê*, usado como clítico-sujeito no PB, deriva da conhecida sequência de erosões fonéticas *cê < ocê < você < vosmecê < vossemecê < Vossa Mercê*. O pronome *usted* do espanhol tem a mesma origem: *Vuestra Merced*, e também passou por uma série de formas reduzidas, até chegar à forma atual. A diferença é que *usted* ainda guarda sua origem de tratamento respeitoso, ao passo que você se tornou no PB o pronome usado para indicar a maior intimidade entre as pessoas.
- embora → de *em boa hora*, uma locução que se transformou em conjunção, advérbio, preposição.
- oxe → a interjeição *oxente!*, característica das variedades nordestinas, deriva de *ó, gente!* com ensurdecimento da palatal /ʒ/. Em muitas ocasiões se reduz a *oxe*, principalmente em final de enunciado.
- pelo amor de → essa expressão do português clássico, já gramaticalizada (a palavra *amor* aí perdeu seu sentido original), se transformou em preposição em diversas variedades do PB. Nesse processo ela sofreu erosão fonética e assumiu diversas formas: *pramode, prumode, pormode, mode* etc.
- sebesta → em variedades linguísticas de Minas Gerais ocorre com frequência a interjeição *sebesta*, que provém do imperativo *deixa de ser besta!*
- sô → o tratamento *senhor*, por causa de seu uso frequente, se reduziu a *sô*, quando empregado como título *sô João, sô Geraldo, sô Pedro*, ou como marcador conversacional em forma de vocativo em final de enunciado: "Vamo lá, sô!" O feminino correspondente é *sá*: *sá Maria, sá Josefa, sá Luzia*. De *senhor* também deriva a forma *sinhô*, com base na qual se formou o feminino *sinhá*. Fenômeno semelhante ocorreu em francês, onde a locução *mon seigneur* ('meu senhor') se reduziu a *monsieur*, pronunciado como [məsiø]. O inglês *sir* [sɐɹ] por seu turno, é redução de *sire* [saiə] que por sua vez vem do francês *sire < sieur < seigneur*. O latim *mea domina* ('minha senhora') está na origem do francês *madame*, que passou para o inglês como *madam*, muitas vezes pronunciado *ma'am*, tratamento que se emprega para falar com a rainha da Inglaterra.
- tá → o verbo *estar* é dos mais empregados na língua, junto com *ser* e *ter*. Além de seu sentido pleno, também é amplamente usado como auxiliar, ou seja, como partícula gramatical, e como marcador conversacional (tag *question* no final de enunciados: "Me espera um minutinho aí, tá?"). Assim, na fala espontânea, seu infinitivo se reduz a *tá*, erosão fonética que também atinge sua conjugação nos tempos usados com maior frequência: *eu tô, cê tá, eu tive, eu tava*. Com isso, ele se torna monossilábico no infinitivo, uma característica dos verbos mais usados na língua (e em todas as línguas do mundo): *ser, ter, dar, ver, vir, ir*. O verbo *fazer*, de amplíssimo emprego, não passou por essa erosão em seu infinitivo, mas as formas conjugadas mais frequentes, *faz, fez, fiz*, são monossilábicas.
- tchau → essa palavra, que nos parece hoje tão brasileira, é de origem italiana: *ciao*. Na Itália, ao contrário do que ocorre no Brasil, ela é usada como saudação no encontro e não na despedida, como entre nós. Sua origem é a palavra *schiao*, do dialeto veneziano, com o significado de 'escravo'. *Schiao* foi o que sobrou de uma oração completa: 'eu sou seu escravo', isto é, 'estou a seus serviços'.
- xovê → de *deixa eu ver*. Essa locução, quase sempre imperativa, é de uso muito frequente em PB, o que ocasionou sua erosão fonética. Coisa semelhante ocorreu em francês, onde a oração *je ne sais pas* ('eu não sei') se reduz, na fala corrente, a [ʃe'pa], assim como *je suis* ('sou' / 'estou') se reduz a [ʃɥi]. Em inglês americano corrente, orações como *I am going to, I have got to, I want to* se reduzem a *I gonna, I gotta, I wanna*.

(a) Na formação do chamado *futuro românico*, a conjugação latina do tempo futuro (*cantabo, cantabis* etc.) foi abandonada em proveito da perífrase composta de *infinitivo + habeo*, do verbo *habere* ('ter'). Assim, no lugar de *cantabo* surgiu *cantare habeo* ('cantar' + 'tenho', isto é, 'tenho de cantar'). Em português, esse uso de *haver* com infinitivo ainda se preserva em expressões como "*hei de vencer*". Essa perífrase se gramaticalizou e, como sempre ocorre nesses processos, sofreu *erosão fonética, enfraquecimento semântico* e *fixação morfossintática*. Em todas as línguas românicas, o futuro passou a ser expresso pela junção do infinitivo e do verbo *habere*:

ITALIANO		FRANCÊS		ESPANHOL		PORTUGUÊS	
AVERE *pres.*	CANTARE *futuro*	AVOIR *pres.*	CHANTER *futuro*	HABER *pres.*	CANTAR *futuro*	HAVER *pres.*	CANTAR *futuro*
ho	canterò	ai	chanterai	he	cantaré	hei	cantarei
hai	canterai	as	chanteras	has	cantarás	hás	cantarás
ha	canterà	a	chantera	ha	cantará	há	cantará
abbiamo	canteremo	avons	chanterons	hemos	cantaremos	h(av)emos	cantaremos
avete	canterete	avez	chanterez	h(ab)éis	cantaréis	h(av)eis	cantareis
hanno	canteranno	ont	chanteront	han	cantarán	hão	cantarão

O mais curioso é que o futuro do latim clássico, *cantabo*, já era resultante de uma gramaticalização do mesmo tipo: *canta* + **bho*, um antigo verbo indo-europeu que significava 'ser, existir, crescer'. Essa formação do futuro românico, resultado de gramaticalização, já se vê em boa medida substituída por uma nova perífrase: a do verbo *ir* seguido de infinitivo — *vou cantar*. No PB, o futuro simples — *cantarei* — raramente se usa na fala espontânea. Na escrita mais monitorada, surgiu uma nova perífrase, formada do verbo *ir* no futuro seguido do infinitivo: *irá começar, iremos fazer*,

irão elaborar. Essa nova perífrase parece representar uma solução de compromisso entre a forma com o verbo *ir* no presente — considerada demasiado "coloquial" — e o futuro simples — considerado demasiado "formal".

(b) Na língua latina não existiam artigos. Em todas as línguas românicas, surgiram essas palavras, essencialmente gramaticais. Os artigos resultam da gramaticalização dos *demonstrativos* latinos. Pelo já mencionado processo de erosão fonética, abrandamento semântico e fixação morfossintática, os demonstrativos, palavras longas e tônicas, se tornaram partículas átonas:

LATIM VULGAR	OCCITANO	FRANCÊS	CATALÃO	ESPANHOL	PORTUGUÊS
masc. sing. ILLU	lo / l'	le / l'	el / lo	el	o
masc. plur. ILLOS	los	les	els	los	os
fem. sing. ILLA	la / l'	la / l'	la / l'	la	a
fem. plur. ILLAS	las	les	les	las	os

Esse processo de gramaticalização é universal, ocorre em todas as línguas onde surgiram artigos: eles sempre provêm de demonstrativos. É o caso do inglês *the*, derivado de um antigo demonstrativo germânico. Em alemão, os artigos *der, die, das* também procedem de antigos demonstrativos. Hoje, por sua vez, servem também como demonstrativos: *der Mann*, 'o homem' / 'esse homem'.

(c) A mesma origem dos artigos é compartilhada pelos pronomes que remetem à não-pessoa. Tais pronomes não existiam em latim. Para completar o quadro de pronomes pessoais, surgiram os pronomes românicos de não-pessoa com base nos demonstrativos latinos, o que explica as semelhanças de forma entre artigos e pronomes:

LATIM VULGAR	FRANCÊS	CATALÃO	ESPANHOL	PORTUGUÊS
masc. sing. ILLE (sujeito)	il	ell / ells	él / ellos	ele / eles
masc. sing. ILLI / ILLUI (ob. ind.)	lui	li / els	le / les	lhe / lhes
masc. sing. ILLU (ob. dir.)	le	el	lo	o / -lo
masc. plur. ILLORU (ob. ind.)	leur	——	——	——
masc. plur. ILLOS (ob. dir.)	les	els	los	os / -los

Esse também é um processo de gramaticalização que se verifica nas mais diferentes línguas do mundo. Ele faz parte de um processo mais amplo: DEMONSTRATIVO > PRONOME DE NÃO-PESSOA > CLÍTICO > FLEXÃO VERBAL. Assim, é possível supor que a terminação -t do latim *amat* ('ele ama') seja o resultado final dessa cadeia de mudanças. Em alguma época remota, um demonstrativo passou por todas as fases, até

se transformar em desinência, morfema preso. O ressurgimento de pronomes de não-pessoa nas línguas românicas demonstra o que os estudiosos da gramaticalização postulam: o caráter cíclico do processo.

(d) Em muitas línguas, a palavra que designa 'homem' se gramaticalizou até se tornar um pronome-sujeito indefinido. No português arcaico, existia *home/homem* com esse sentido, como nos exemplos literários abaixo (citados por Said Ali, 1964: 116):

(10) *Em aquel tempo nom podia homem achar em todo o regno de logres donzel tam fremoso nem tam bem feito.*
(11) *Eu te farei taes cousas quaes nunca homẽ fez a seu inimigo.*
(12) *Era tam esprandecente que bem se podia homem veer en ella como em espelho.*
(13) *Certo he grande erro não conhecer homem seu erro.*
(14) *Eu perdi a mor ventura que homem nunca perdeo.*

Essa também é a origem do pronome indefinido francês on, do alemão man, do islandês maður, todos derivados de 'homem'.

(e) Do substantivo latino mente- ('mente') se formou, nas línguas românicas, o sufixo -mente, usado na formação de advérbios de modo. Em latim vulgar, fazer algo *bona mente* significava originalmente fazer algo 'com boa disposição, de boa vontade', fazer algo *forte mente*, 'com decisão de espírito'. No processo de gramaticalização, a palavra perdeu sua autonomia lexical e se transformou num morfema preso. Essa origem explica por que, na formação desses advérbios, a forma do adjetivo deve estar no feminino, já que a palavra mente é feminina. Também esclarece por que, nas sequências de advérbios de modo, é possível eliminar o sufixo -mente, que só aparece no último advérbio da série: "Copacabana, Copacabana, louca, total e completamente louca" (canção "Joia", de Caetano Veloso).

Evolução semelhante se encontra no sufixo -ly do inglês, usado para formar advérbios de modo, assim como seu correspondente alemão -lich: os dois provêm de um antigo adjetivo germânico aparentado à palavra like do inglês moderno, que significa 'semelhante', 'parecido'.

(f) No português arcaico se usava a palavra passo como reforçador da negação: *Não conheço passo disso*. Esse uso permanece até hoje no francês, em que a partícula pas (que também significa 'passo') se tornou obrigatória na expressão da negação. Por meio da metonímia, a palavra pas, de sentido positivo, absorveu a carga semântica negativa de ne ('não'), de modo que no francês falado contemporâneo, a partícula ne é frequentemente omitida: *Je connais pas ton mari* ('não conheço o teu marido').

(g) A palavra latina casa, substantivo, se transformou em francês numa preposição chez, de valor exclusivamente gramatical. Chez tem várias utilizações: *chez moi*

('na minha casa', 'no lugar onde eu moro'); *chez Renault* ('na empresa Renault'); *une maladie rare chez les enfants* ('uma doença rara entre as crianças'); *chez Proust* ('na obra de Proust') etc.

(h) O substantivo hora se transformou no advérbio-conjunção-interjeição ora. Essa também é a origem do francês or, usado como advérbio e conjunção.

(i) Conhecidíssimo é o caso do pronome você. O tratamento respeitoso Vossa Mercê foi sofrendo enfraquecimento semântico, extensão pragmática, erosão fonética e fixação morfossintática, passando por diversos estágios — vossemecê, vosmecê, vassuncê, vancê etc. — até se transformar em você. No PB, esse pronome já alcançou o estágio da cliticização: quando exerce a função de sujeito, e somente nessa função, se reduziu a cê, que ocorre exclusivamente em próclise ao verbo: *cê viu, cê sabe, cê quer?* Em algumas variedades, com destaque para Minas Gerais, é muito corrente a forma ocê, que se contrai com as preposições formando docê, nocê, concê, procê. Na função de sujeito, ocê concorre com cê. Tendo se tornado um verdadeiro clítico-sujeito, não tem nenhuma justificativa a persistência da gramática normativa em classificar você como "pronome de tratamento", já que ele é, de fato, o índice de segunda pessoa mais empregado em todo o Brasil.

Os processos de gramaticalização que implicaram em mudança de categoria gramatical estão bem registrados na aparência superficial de muitas palavras que têm emprego gramatical exclusivo e que revelam as etapas de aglutinação de palavras anteriormente soltas: portanto, porquanto, porventura, contanto, contudo, todavia, talvez, entretanto, enquanto, embora, outrora, agora, entanto, depois, após, dentre etc. Nosso advérbio hoje provém de hodie, que já era uma gramaticalização em latim de *hoc die*, 'este dia', enquanto ontem provém de *ad noctem*, 'à noite'. Verbos (e até sentenças inteiras) que se transformaram em palavras e locuções gramaticais se revelam em qualquer, sequer, seja... seja, exceto, salvo, durante, mediante, não obstante, tomara, pudera, bastante, seja o que for, o que quer que, onde quer que, haja vista etc. Em francês os pronomes indefinidos n'importe quoi, n'importe qui, n'importe où, n'importe comment etc. se formaram com base no verbo importer e devem ser traduzidos por 'qualquer coisa', 'qualquer um', 'qualquer lugar', 'de qualquer forma' etc. Outros pronomes indefinidos franceses têm origem em gramaticalizações como quelque chose ('algo') e quelqu'un ('alguém'). O mesmo vale para os nossos indefinidos algum, alguém, nenhum etc. O francês peut-être, literalmente 'pode ser', mas com o sentido do nosso talvez, é em tudo semelhante ao inglês maybe (de *may*, 'poder', e *be*, 'ser') e ao catalão potser, 'talvez'. Em inglês também são muitas as palavras gramaticais provenientes de aglutinações: whenever, whoever, wherever, forever, whatsoever, notwithstanding, tonight, today, together, everyday, nowadays, afternoon, however, anyhow, none, someone, sometimes, somewhere, alone (de *all*, 'todo', e *one*, 'um': 'sozinho', que em francês se traduz exatamente por tout seul, literalmente 'todo só') etc. A palavra 'coisa' também aparece com frequência em palavras gramaticais: quelque chose, em francês, com o sentido de 'algo'; cosa, em italiano, como interrogativo com o sentido de 'o quê?'; qualcosa, também em italiano, 'algo'; anything, everything, nothing, something em inglês, todos formados com base em *thing*, 'coisa' etc.

A metáfora também incide fortemente no uso de termos que designam o corpo humano e suas partes para formar itens gramaticais. Uma vez que o corpo é nosso primeiro e principal meio de contato físico com o mundo exterior, não admira que os nomes de suas partes acabem por se transformar em conceitos abstratos. O inglês *back*, que designa as costas, o dorso, e também 'para trás', 'de volta', é o caso mais notável entre as línguas mais difundidas. O francês *chef*, derivado do latim *caput*, 'cabeça', também se enquadra nesse caso. Nossa palavra *fronte* deriva do latim *fronte-*, 'testa', e com ela formamos palavras como *defronte, confronto, confrontar, afronta* etc. Do espanhol *fruente* > *frente*, também derivado do latim *fronte-*, herdamos *frente*, de onde vêm *em frente a, de frente a, enfrentar* etc. Em galego a locução *cara a* indica 'na direção de', 'com relação a', e provém da palavra *cara*, 'rosto, face'. Os sentidos metafóricos de palavras como *face, olho, cabeça, pé, dente, boca, nariz, cara, mão, perna, dedo* etc. são praticamente incontáveis e ocorrem em todas as línguas humanas. Um rápido apelo à memória nos traria *boca do monte, pé de serra, olho do furacão, olho d'água, braço de rio, braça (medida), boca de rio, face da montanha, flanco da montanha, costa* ('litoral'), *polegada, pé* (medida), *côvado* (medida, derivado de *cúbito*, 'cotovelo'), *palmo* (medida), *dedo* (medida), *colo da montanha, encarar, pé de* (árvore frutífera) etc. Os verbos *expedir, despedir e impedir* têm o radical *ped-*, do latim *pede-* ("pé"). Para *mão* e *cabeça* os dicionários trazem sempre dezenas de significados e acepções. Como dois terços do que apreendemos do mundo nos entram pelos olhos, o verbo *ver* sempre está associado, nas línguas humanas, com a ideia de 'conhecimento', 'sabedoria', 'ideia' etc. [▶278].

A metaforização das partes do corpo segue uma direção constante, sempre do mais concreto para o mais abstrato:

PESSOA → OBJETO → ESPAÇO → TEMPO → PROCESSO → QUALIDADE

Examinando essa sequência, percebemos que, de fato, ela é muito produtiva nos fenômenos de gramaticalização. Como exemplo, basta pensar nas partes do corpo humano (= PESSOA) que passaram a designar OBJETOS inanimados (*cabeça* de alfinete, *dente* da engrenagem, *boca* do canhão, *pé* da mesa etc.), para daí designar localizações ESPACIAIS (*pé* da serra, *boca* do rio, *costa, braço* de mar etc.), chegando à indicação de TEMPO (*boca* da noite), de PROCESSO (fazer algo com um *pé nas costas, nas coxas* ou *de olhos fechados*, ficar *de olho*, de *nariz empinado*, empurrar *com a barriga*, tocar de *ouvido, de corpo mole*, saber de *cor* [= coração], com *a boca no mundo*)

e de QUALIDADE (filme *cabeça*, música *bunda*, indivíduo *pentelho*, comportamento *escroto*). Na sequência ESPAÇO > TEMPO podemos retomar o exemplo do inglês *back*, que significa 'voltar para um lugar' e também 'voltar no tempo'.

As partes do corpo se revelam grandes produtoras de metáforas gramaticalizadas nas mais diferentes línguas do mundo (Heine e Kuteva, 2002):

- ewe:
  i. *épé megbé fá*
  dele costas é.frio
  'As costas dele são frias'
  ii. *é-le xɔ-á megbé*
  ele.é casa-a costas
  'Ele está de volta da casa'
- hebraico moderno:
  i. *pney ha-yeled*
  face (de) a criança
  'A face da criança'
  ii. *hi omedet li-pney ha-bayit*
  ela fica de-face (de) a-casa
  'Ela fica de frente para a casa'
- albanês:
  në bark të javës
  em barriga-a semana
  'na barriga da semana' = 'no meio da semana'
- húngaro:
  bel-ügy-minisztérium
  intestinos-assuntos-ministério
  'Ministério do Interior'
- náhuatl (México)
  huēi ālтepētl i-yōllò-co
  grande cidade seu-coração-em
  'Na grande cidade'
- galês (País de Gales)
  ger fy mron
  perto meu peito
  'Diante de mim'

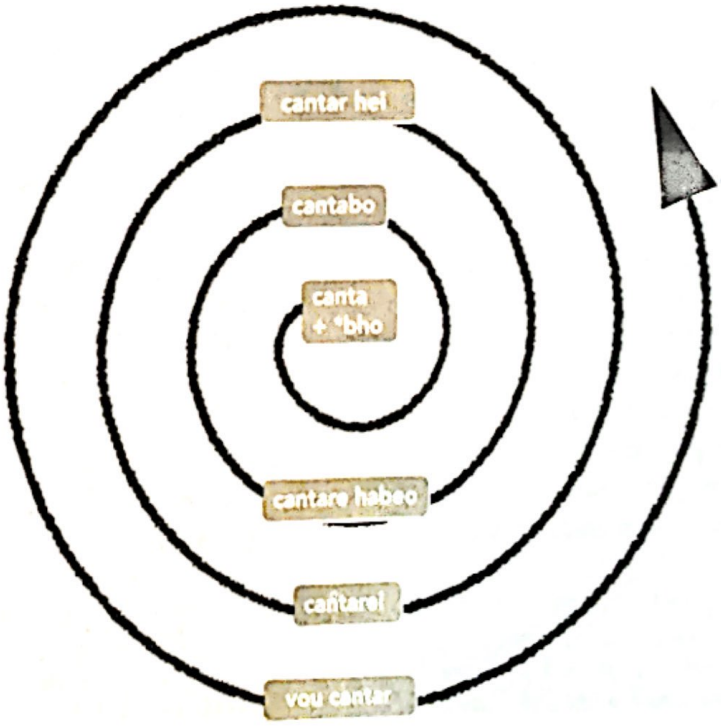
A palavra para 'corpo' deu origem, em muitas línguas, a um pronome reflexivo, como se disséssemos 'eu levantei meu corpo às 8 horas', para expressar *eu me levantei às 8 horas*. Isso ocorre nas línguas vai, ibibio, didinga, bagirmi, shilluk, lango, päri, lele, !xun, iorubá, moru, luo, gidar, xdi, margi entre outras. Um exemplo do iorubá, língua trazida para o Brasil por muitos escravos:

- iorubá (Nigéria, Benin):
  Nwosu rí ara rẹ̀.
  Nwosu viu corpo seu
  ('Nuwosu se viu')
- moré (Burkina-Faso):
  a kũ a mega.
  ele matar seu corpo
  ('Ele se matou')

Em inglês, existem os pronomes *anyboby, somebody, nobody, everybody*, todos derivados de *body*, 'corpo'.

Uma das palavras mais polissêmicas da língua é ponto. Basta abrir qualquer dicionário e ver que as acepções se sucedem às dezenas (o dicionário *Houaiss* oferece 54 delas, fora outras tantas locuções). Essa palavra deriva de punctu-, particípio passado do verbo pungĕre, cujo primeiro sentido era 'picar', como faz um inseto. Um *punctu-* portanto era, em latim, inicialmente, uma picada de inseto. Ora, como a picada deixa na pele uma pequenina marca circular, a partir desse elemento concreto, físico, a palavra *ponto* alçou seu voo metafórico e foi adquirindo, ao longo do tempo e nas diferentes línguas em que é usada, muitos e muitos sentidos abstratos.

Os processos de mudança induzidos pela gramaticalização também ocorrem no campo da sintaxe das línguas. Aqui o principal fator cognitivo em ação é a reanálise sintática. Os falantes passam a interpretar de forma nova o encadeamento dos constituintes no eixo sintagmático. O caso do verbo *ir* que se transformou em expressão do futuro é exemplar [▶617]. De início, *vou visitar Marlene* era analisado como [*vou*] [*visitar Marlene*], ou seja, 'estou me deslocando agora no espaço' tendo como meta de chegada 'visitar Marlene'. No entanto, como já vimos, todo deslocamento espacial é ao mesmo tempo um deslocamento temporal e, a partir dessa realidade do mundo empírico, foi possível reanalisar a sentença como [*vou visitar*] [*Marlene*], ou seja, 'em algum momento por vir, me deslocarei no espaço para visitar Marlene'. A impregnação da metáfora temporal se tornou tão profunda e arraigada que, a noção de deslocamento espacial desapareceu por completo em expressões como *vou dormir, vou ficar, vou fechar os olhos e fingir que você não está aqui*... Com isso, o verbo *ir* se esvaziou semanticamente e se tornou um item gramatical usado para expressão do futuro.



4.4.3 A *analogia*

O terceiro e último fator cognitivo que vamos examinar como grande provocador da mudança linguística é a analogia. A *analogia* é um processo cognitivo por meio do qual os falantes da língua tendem a regularizar formas irregulares e menos gerais com base em outras formas, mais regulares e de emprego mais frequente. No estudo diacrônico de qualquer língua, é possível observar a analogia em ação, na medida em que formas outrora irregulares passaram a se enquadrar em paradigmas regulares. Não por acaso, as formas irregulares que sobrevivem por mais tempo são precisamente aquelas mais usadas e usadas com maior frequência e que, justamente por isso, resistem aos processos de regularização paradigmática. É o caso, por exemplo, dos verbos mais empregados em PB (e em todas as línguas): *ser, ter, dar, ir, ver* etc. O próprio fato de serem palavras curtas demonstra sua alta frequência de uso, pois quanto mais uma palavra é usada (sobretudo quando passa pelo processo de gramaticalização) mais tendência ela apresenta de, ao longo do tempo, se contrair (cf. *Vossa Mercê* > *vossemecê* > *vosmecê* > *você* > *ocê* > *cê*).

O linguista Guy Deutscher, que já citamos ao tratar da metáfora, assim se expressa acerca da analogia (2005: 174):

> Como se lembrará qualquer um que já tenha tentado aprender uma língua estrangeira, quanto mais ordem e regularidade se puder apreender, menos formas individuais será preciso memorizar. [...] Se não fosse possível extrair padrões recorrentes da massa de informações novas a ser absorvida, nossas mentes simplesmente se atolariam nos detalhes.
>
> A capacidade de depreender padrões não é crucial apenas quando se aprende uma língua estrangeira: também é vital para as criancinhas às voltas com sua língua materna. Os bebês não sugam a língua materna junto com o leite materno, eles precisam depreender por conta própria todo o bendito sistema, e a massa de informação que eles têm de assimilar é de enlouquecer. O fardo fica mais leve, porém, quanto mais padrões recorrentes eles puderem identificar. Por isso não admira que as crianças ajam na suposição de que quase tudo na língua deve seguir regras simples e regulares, e daí vêm os erros mimosos do tipo "um afoto", "se eu sesse", "mais grande" etc. Esses equívocos nada mais são do que tentativas perfeitamente sensatas de introduzir ordem em recantos da língua que se mostram um tanto bagunçados e irregulares. Às vezes, as crianças conseguem até mesmo passar a perna no princípio básico da arbitrariedade do signo. Não contentes com a ideia de que as palavras significam alguma coisa somente por convenção, elas encontram padrões significativos nas palavras mais aleatórias. Um caso muito citado é o de um espertinho de fraldas que, quando apresentado a um garfo (*fork*) com apenas três dentes, estudou o objeto atentamente e, com a maior naturalidade, decretou que se tratava de um *threek*[3].
>
> À medida que vão crescendo, as crianças gradualmente aprendem que há áreas de sua língua que não se submetem a regras regulares, de modo que a maioria dos equívocos é corrigida: "um afoto" é substituído por *uma foto*, "se eu sesse" por *se eu fosse*, "mais grande" por *maior*

[3] A criança deduziu que *fork* provinha de *four* ('quatro') e, assim, ao ver um garfo de três dentes, formou a palavra *threek*, a partir de *three* ('três').

> etc. Mesmo assim, se tais erros persistirem para além da infância, é possível que ganhem terreno e acabem sobrepujando formas bem estabelecidas. Em inglês antigo, por exemplo, os nomes *eye* e *cow* tinham plurais irregulares: *eyn* e *kine*. Mas, em algum momento, os "erros" *eyes* e *cows* pegaram e acabaram por usurpar o lugar das formas originais.

De fato, as crianças brasileiras, enquanto vão adquirindo a língua, deduzem padrões regulares que vão tentando aplicar a tudo o que ouvem. Daí as formas, tipicamente infantis, *eu fazi, eu di, eu sabo, se eu queresse, se eu sesse, eu pido* etc., todas perfeitamente analógicas. Conheço uma menina que, quando bem pequena, passou por um lugar onde havia muitos búfalos. Ao saber o nome do animal ela disse: *"Quantos búfalos! E quantas vácalas!"* E outra que dizia *rezário* por *rosário*, *vaso solitário* por *vaso sanitário* e *Bico do Jaraguá* por *Pico do Jaraguá*. De um menino de 5 anos ouvi certa vez dizer que morava numa casa junto da *Mata Plântica*, por *Atlântica*. As analogias infantis operam em todos os níveis da língua: fonético, morfológico, sintático, semântico.

Um bom exemplo de analogia que prosperou em português foi o que se deu com os verbos *impedir, expedir, despedir*, que, ainda no século XVII, se conjugavam *eu impido, eu expido, eu despido* etc. (como em espanhol até hoje). No entanto, por analogia com a conjugação de *pedir* (*eu peço*), aqueles verbos, que não têm nenhum parentesco etimológico com *pedir*, passaram a se conjugar como ele. Já vimos que os verbos *impedir, expedir, despedir* têm em sua etimologia a raiz *ped-* ('pé'): *impedir* seria travar o pé de alguém, não deixar que se vá. Já o verbo *pedir* provém de *petire* (< *petĕre*).

Desde o surgimento da linguística histórica, no século XIX, a analogia vem sendo usada como explicação para a ocorrência, na língua, de formas que se desviam do resultado esperado da aplicação das chamadas "leis fonéticas". Numa visão bastante positivista, os pesquisadores alemães chamados *neogramáticos*, que se debruçaram intensamente sobre as questões de mudança linguística na segunda metade do século XIX, atribuíam à analogia todo e qualquer resultado de mudança que não "obedecesse" as "leis" que eles tinham estabelecido.

Depois de abandonada como explicação para a mudança, a analogia voltou a ser considerada, agora sob nova ótica, porque é evidente que, como processo cognitivo, ela de fato atua na produção de formas linguísticas inovadoras que, uma vez aceitas pela comunidade de falantes, acabam por substituir as formas antigas.

Segundo Saussure, a analogia pode ser vista como resultado de um quarto termo proporcional, numa equação do tipo:

A : B : : C : x → 'A está para B, assim como C está para x'

Assim, por exemplo, em muitas variedades do PB, ocorre o infinitivo *ponhar* no lugar do clássico *pôr* — a regra aqui, portanto é: *sonho* está para *ponho*, assim como *sonhar* está para *ponhar*. Com isso, o verbo *pôr*, com seu infinitivo peculiar, altamente irregular, se torna regular. Evidentemente, não há em nada disso um pensamento racio-

nal, algébrico, consciente da criação de uma forma nova — trata-se, mais uma vez, de um processamento cognitivo que apela para a sempre poderosa intuição linguística de qualquer falante normal. Essa tendência a criar formas novas e regulares por meio da analogia é conhecida como princípio do isomorfismo: 'Para cada significado deve haver só uma forma linguística'. No caso de *ponhar*, a ação do princípio do isomorfismo é nítida: em português, os infinitivos verbais podem terminar em *-ar*, *-er* e *-ir*, de modo que, ao buscar um infinitivo para *eu ponho*, os falantes deduzem *ponhar*. Três terminações diferentes para representar o infinitivo constituem, por outro lado, um excesso, tanto que todo verbo novo criado na língua se encaixa na primeira conjugação, *-ar*. Basta pensar nos verbos surgidos com as inovações tecnológicas: *telefonar*, *televisar*, *telegrafar*, *automatizar*, *robotizar*, *liquidificar*, *deletar*, *customizar*, *visualizar*, *clonar*, entre tantos outros, são todos da primeira conjugação. Com base em empréstimos tomados de outras línguas, também são criados novos verbos, sempre da primeira conjugação: *printar*, *startar*, *palletizar*, *shrinkar*, *roughear* são verbos empregados na indústria gráfica e editorial, todos originados de palavras vindas do inglês.

Muitas formas linguísticas usuais no PB se devem a processamentos analógicos:

1. Palavras são conduzidas a um novo *gênero* gramatical: *o ferrugem*, *a tomate*, *o couve*, *o omelete*, *a dó*, *o alface*, *a sabiá*, *o hortelã* etc. Nesses casos, a ausência das vogais átonas finais *-o* e *-a* que se tornaram, na história da língua, morfemas de gênero masculino e feminino, respectivamente, deixa liberdade ao falante para a atribuição de gênero. No caso de *grama* ('unidade de medida'), que no PB é exclusivamente do gênero feminino, a analogia se deu com todas as incontáveis palavras terminadas em *-a* que são do gênero feminino. Por ser uma palavra usada com altíssima frequência, foi conduzida pelos falantes à lista das palavras femininas do léxico da língua. Observe que outras palavras de origem grega terminadas em *-ama*, mas de uso bem mais restrito, não sofreram o processo analógico: *o programa*, *o telegrama*, *o panorama*, *o anagrama* etc. A manutenção do gênero masculino de *programa* se deve provavelmente ao intenso emprego da palavra nos meios de comunicação (*programa de rádio*, *programa de televisão*), que contribuíram para a ampla difusão do termo e de seu respectivo gênero. A palavra *grama*, no entanto, usada há séculos na língua, se tornou feminina. Falantes de variedades altamente estigmatizadas costumam flexionar o advérbio *menos*, interpretado como adjetivo, gerando a forma *menas*: *menas gente*. Em áreas do Nordeste, a palavra *tapa* é considerada feminina — *vou te dar uma tapa na cara* —, por analogia com as demais palavras terminadas em *-a*.

2. Palavras são conduzidas a um novo número gramatical: é o caso conhecido de *óculos* que no PB é exclusivamente singular, decerto por analogia com *pires*, *ônibus*, *lápis*, e também pela evidência empírica de que se trata de um objeto

único e não de um par. Ninguém usa um *óculo*, mas sempre um objeto único com duas lentes. Essa evidência empírica da unidade do objeto também atingiu palavras que, no português clássico, eram usadas sempre no plural: *tesouras, calças, ceroulas, tenazes*, e que no PB só são empregadas no singular.

3. Palavras como *chapéu* e *troféu* costumam ser pluralizadas como *chapéis* e *troféis*, por analogia com *papel/papéis, anel/anéis, pincel/pincéis* etc. Como no PB o [l] final se vocalizou completamente na maioria das variedades em /ʊ/, a pronúncia das palavras terminadas em *-éu* e *-el* se tornou idêntica. Sendo as palavras terminadas em *-éu* muito poucas no uso diário, se comparadas às terminadas em *-el*, os falantes as incluíram na classe dos plurais em *-éis*.

4. Regências verbais sofrem alteração: o verbo *assistir*, por sua identidade semântica com verbos transitivos diretos como *ver, presenciar, frequentar, observar* etc. se tornou exclusivamente transitivo direto no PB [▸523]. O verbo *buscar* vem sendo usado com cada vez mais frequência, sobretudo nos meios de comunicação, como transitivo indireto, regendo a preposição *por*: *Buscando por melhores salários*. Esse uso pode ser analógico ao do verbo *procurar*, de sentido semelhante, que pode se construir com a preposição *por*: *Tem gente procurando por você lá fora*. O verbo *permitir* é interpretado como um *causativo* [▸535], à semelhança de *deixar* e, por isso, é empregado como transitivo direto: *Isso permite eu ganhar mais tempo*. Em muitos gêneros textuais escritos, ele ocorre com os clíticos acusativos de não-pessoa: *Isso o permitiu levar adiante o projeto*.

5. Palavras são consideradas, por equívoco, como derivadas de outras, com as quais não têm relação. Por exemplo, é comum a inserção de um c em *aficcionado* por uma falsa etimologia com *ficção*. O vocábulo no entanto provém diretamente do espanhol *aficionado* e tem a mesma origem do português *afeição*.

6. Os nomes dos meses *setembro, outubro, novembro* e *dezembro* ganharam um *-o* final por analogia com *janeiro, fevereiro, março, maio, junho, julho*. Pelas mudanças regulares da língua, era de se esperar que terminassem em *-e* (como em espanhol: *setiembre, octubre, noviembre, diciembre*).

7. Analogias por hipercorreção também ocorrem, criando dessa vez formas irregulares. O particípio passado *pego* já foi considerado próprio de "ignorantes" e hoje é usado pelas pessoas mais letradas. O mesmo processo que gerou *pego* gerou também *trago* (do verbo *trazer*) e *chego* (do verbo *chegar*), ainda muito estigmatizados pelos falantes escolarizados. Esses particípios irregulares decerto se formaram por analogia com *pago*, do verbo *pagar*, e com outros particípios irregulares (*entregue, aceito, feito, dito* etc.). Por outro lado, na história da língua, uma enorme quantidade de particípios latinos foram refeitos de acordo com as terminações regulares da maioria dos verbos. Compare alguns deles com os particípios dos mesmos verbos em italiano padrão:

LATIM > PORTUGUÊS	ITALIANO
◗ *míttere, missu-* > meter, metido	(*méttere, messo*)
◗ *mórdere, morsu-* > morder, mordido	(*mórdere, morso*)
◗ *offéndere, offensu-* > ofender, ofendido	(*offéndere, offeso*)
◗ *offerire, offertu-* > oferecer, oferecido	(*offrire, offerto*)
◗ *opprímere, oppressu-* > oprimir, oprimido ◗ *rúmpere, ruptu-* > romper, rompido	(*opprímere, oppresso*) (*rómpere, rotto*)
◗ *assúmire, assumptu-* > assumir, assumido	(*assúmere, assunto*)
◗ *téndere, tensu-* > tender, tendido	(*téndere, teso*)
◗ *vincere, victu-* > vencer, vencido	(*víncere, vinto*)
◗ *suspéndere, suspensu-* > suspender, suspendido	(*sospéndere, sospeso*)

Alguns exemplos de mudança fonética ocasionada por analogia:

(a) Do francês *camion* o esperado seria *camião*, mas por analogia com *caminho* surgiu *caminhão*.
(b) De *frio* mais o sufixo *-ento* se esperaria *friento*, mas por analogia com *calorento* temos *friorento*.
(c) De *herba* mais o sufixo *-izar* teríamos *herbizar*, mas por analogia com *arborizar* o que temos é *herborizar*.
(d) Do francês *cheminée* teríamos *cheminé*, mas por analogia com *chama*, temos *chaminé*.
(e) A palavra *fome* provém do latim *fame* (donde *faminto*). Para explicar o *o*, alguns autores recorrem à hipótese de uma analogia com a forma *come* do verbo *comer*.
(f) O numeral *noventa* provém de *nonaginta*. O *v* se deve à analogia com o numeral *nove*.
(g) A palavra *floresta* provém do latim *forestis* 'bosque externo', mata que fica *fora* da cidade. Aqui a analogia se deu com a palavra *flor*.
(h) Pelas mudanças fonéticas regulares, o latim *fabulare* deveria ter resultado em *fabrar*. A forma *falar* é explicada por analogia com *calar*, verbo com o qual *falar* compõe um par muito usado.

No campo da morfologia, é possível detectar inúmeros processos de mudança decorrentes da analogia. De fato, desde os filósofos gregos, a analogia vem sendo percebida como a busca da regularidade dos paradigmas. Em línguas como as românicas, que têm uma complexa rede de flexões, o impulso analógico opera quase sempre com o objetivo da regularização.

Na sintaxe, a analogia também atua, como vimos com as regências de alguns verbos. As construções mais frequentes em PB com o verbo *preferir* — *prefiro [mais] X do que Y* — se devem por analogia a *gostar, apreciar*, em que a comparação é frequente. A sintaxe do verbo *pedir* em PB — *ele pediu para eu vir* — é análoga à do verbo *dizer*: *Ele disse para eu vir*. A regência *namorar com* — *Paula namora com Gustavo* — é análogica à de *noivar com, casar com*.

Por fim, convém ressaltar que os três fatores de mudança linguística examinados aqui — economia linguística, gramaticalização, analogia — interagem o tempo todo numa rede complexa de processamentos cognitivos, de modo que não é possível delimitar com exatidão o que se deve a cada um deles: é mais produtivo encarar as inovações linguísticas como o resultado dessa rede complexa de cognição que, decerto, inclui outros processamentos.

Por exemplo, a simplificação dos paradigmas verbais apresentada acima é consequência da economia linguística, como analisamos, mas seu resultado é uma gramaticalização dos pronomes-sujeitos, que se transformam em clíticos, e uma regularização do paradigma de conjugação, ou seja, a criação de modelos analógicos.

4.5 A efetivação da mudança linguística

4.5.1 De baixo para cima

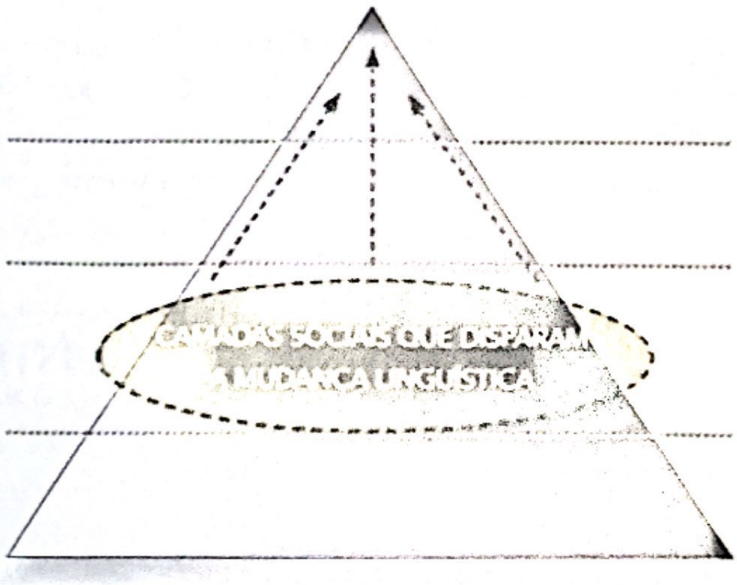
Os estudos clássicos da sociolinguística variacionista nos anos 1960-1970 mostraram que as mudanças linguísticas mais frequentes são aquelas que partem das camadas médias baixas da população e vão se incorporando progressivamente na fala das camadas médias altas, até atingir as camadas sociais mais altas. Uma vez instaladas nesse grupo social dominante, as formas inovadoras deixam de sofrer avaliação negativa, ou seja, deixam de ser vistas como "erro" e se tornam plenamente aceitas — exceto, é claro, pelo reduzido número dos "puristas maníacos", como os classifica o dicionário *Aurélio* no verbete *aperceber*.

As mudanças que surgem na fala das camadas mais baixas, sobretudo (no caso brasileiro) entre falantes rurais e rurbanos (aqueles que habitam as periferias e comunidades pobres das grandes cidades e conservam traços da cultura rural), são fortemente rejeitadas pelos falantes das classes médias e altas, justamente por identificarem os cidadãos mais desfavorecidos. Além disso, essas mudanças são mais radicais, ou seja, levam os processos sociocognitivos que estudamos acima mais adiante e de forma mais acelerada, o que causa maior rejeição por parte dos mais letrados. Essas inovações são classificadas como traços descontínuos, porque não avançam para cima no espectro das classes sociais. Já as inovações surgidas nas camadas médias baixas constituem traços graduais, na medida em que ocorrem na fala de toda a população, diferenciando-se seu uso pela frequência: as camadas médias altas e altas usam com menos frequência essas formas inovadoras, sobretudo em estilos de fala (e principalmente de escrita) mais monitorados.

Dessa forma, inovações como a deslateralização da consoante [ʎ] (*trabalho* > *trabaio*), a rotacização nos encontros consonantais com [l] (*placa* > *praca*), a redução

do paradigma verbal a duas formas (*eu canto / tu-você(s)-ele(s)-nós canta*) etc. são altamente estigmatizadas pelos falantes mais escolarizados.

O particípio irregular *pego* recebia avaliação negativa num dicionário produzido na década de 1950, enquanto em 1999 outro dicionário o apresentava de maneira neutra, demonstrando a plena aceitação da forma nova por parte dos falantes das VUP.



Quando a forma inovadora se instala definitivamente no repertório linguístico dos falantes das VUP, a avaliação social muda inteiramente. Assim, se confrontados com o fato de que essas inovações não têm respaldo na TGP, a reação mais comum costuma ser: "Pode até estar errado, mas é assim que todo mundo fala". Essa mesma condescendência, no entanto, não é empregada para qualificar as inovações presentes nas VLE: aqui elas são pura e simplesmente "erros", cometidos por "gente ignorante".

4.5.2 Coisas da juventude

As mesmas investigações sociolinguísticas nos revelam que é na língua falada pelas gerações mais jovens que podemos localizar indícios bem claros da mudança linguística. E é também justamente por isso que as pessoas mais velhas costumam avaliar negativamente a fala dos mais jovens, dizendo que eles têm "vocabulário pobre", "só usam gíria" ou "falam muito errado". Essas acusações aparecem em todas as culturas, desde a mais remota antiguidade.

Quando temos a chance de comparar e analisar a língua falada por duas ou três gerações distintas de falantes (digamos, um avô de 75 anos, seu filho de 45 e seu neto de 15), encontramos exemplos concretos de mudança em todos os níveis da língua na passagem de uma geração para outra (os sociolinguistas chamam isso de mudança em tempo aparente [▶122]). Isso porque, como revelou também a sociolinguística, nós não falamos a língua exatamente do jeito como a aprendemos com nossos pais,

mas sim como aprendemos, na fase da pré-adolescência, com as pessoas da nossa faixa etária ou alguns anos mais velhas que nós. Vimos acima o exemplo do uso de *este/esse*, com gradual desaparecimento de *este* na fala das gerações mais jovens.

Outro fenômeno importante para a compreensão da gramática do PB é a retomada anafórica de objeto de não-pessoa, como vamos ver em [▸596]. Investigando esse fenômeno, Duarte (1989: 27) nos mostra o processo de mudança de acordo com a faixa etária. Segundo a autora, existem quatro possibilidades dessa retomada anafórica:

1) clítico → Deixei **o envelope** no quarto, você **o** pega para mim?
2) pronome lexical → Deixei **o envelope** no quarto, você pega **ele** para mim?
3) sintagma nominal → Deixei **o envelope** no quarto, você pega **o envelope** para mim?
4) categoria vazia → Deixei **o envelope** no quarto, você pega [ø] para mim?

FAIXA ETÁRIA	CLÍTICO	PRON. LEX.	SINT. NOM.	CAT. VAZIA
15 a 17 anos	0 %	23,5 %	10,7 %	65,8 %
22 a 33 anos	5,0 %	20,7 %	13,5 %	60,8 %
34 a 46 anos	3,0 %	16,1 %	13,3 %	67,6 %
acima de 46 anos	5,7 %	15,1 %	18,0 %	61,2 %

A ausência total de clíticos na fala do grupo mais jovem nos informa que esses pronomes não pertencem ao VGB e que, quando são empregados, sua ocorrência está estreitamente vinculada à escolarização, pois só as pessoas com acesso à norma linguística veiculada pelo ensino conhecem esses pronomes — e mesmo assim, como os dados revelam, usam eles muito raramente. Outra informação importante é que a categoria vazia (também chamada pronome nulo) representa, de longe, a estratégia anafórica mais empregada pelos falantes. Esses dados são fundamentais para que se possa traçar estratégias de ensino da língua que levem em conta a realidade do PB e proporcionem de forma mais eficiente o acesso dos aprendizes às regras linguísticas que, mesmo não fazendo parte do vernáculo, têm uso garantido nos GTM.

4.5.3 Da fala para a escrita

Outra forma de detectar a instalação definitiva de uma inovação linguística é analisando seu avanço gradual desde os gêneros orais menos monitorados até os gêneros escritos mais monitorados.

Até cinquenta anos atrás, por exemplo, colocações pronominais como essas

1. O diretor poderia me ajudar.
2. O diretor poderá me ajudar.
3. Se o diretor tivesse me ajudado...
4. Assim, vamos nos tornando mais independentes...
5. Aos poucos temos nos tornado mais independentes...

eram consideradas "viciosas" e "erradas" — as formas "corretas" deveriam trazer o clítico *antes* do verbo auxiliar ou, quando possível, em *mesóclise*:

6. O diretor me poderia ajudar / O diretor poder-me-ia ajudar.
7. O diretor me poderá ajudar / O diretor poder-me-á ajudar.
8. Se o diretor me tivesse ajudado...
9. Assim, vamo-nos tornando... /... nos vamos tornando... /... vamos tornando-nos...
10. Aos poucos nos temos tornado /... temo-nos tornado...

No entanto, sob a pressão da gramática do PB, em que o clítico ocorre sempre em próclise ao verbo principal, as regras da TGP foram sendo abandonadas em favor das construções antigamente tidas por "erradas". Hoje em dia, só em textos escritos altamente rebuscados encontramos as colocações pronominais dos exemplos 6-10. Na escrita jornalística, sobretudo, elas praticamente não ocorrem. Aliás, os manuais de redação das grandes empresas jornalísticas já baniram a mesóclise de suas prescrições.

Desastrosamente, a maioria dos livros didáticos disponíveis no mercado ainda tentam prescrever colocações pronominais que não são seguidas por quase ninguém, nem sequer os escritores contemporâneos, há pelo menos meio século:

> O pronome nunca faz ênclise com os futuros. Jamais se dirá: *Ela teria-me dito a verdade.* (Colocações possíveis: *Ela me teria dito a verdade* ou *Ela ter-me-ia dito a verdade.*) [Faraco / Moura, 2008, vol. 8: 205]

4.5.4 Será que é mesmo assim?

A sociolinguística variacionista surgiu nos anos 1960 nos Estados Unidos, por impulso dos trabalhos pioneiros de William Labov, considerado até hoje o maior nome dessa disciplina. No entanto, daquela época para cá — meio século simplesmente —, ocorreram profundas transformações sociais e culturais mundo afora, com destaque para o desenvolvimento das novas tecnologias de informação e de comunicação. A delimitação precisa, por exemplo, entre o que é falado e o que é escrito se torna cada dia mais difícil, senão impossível — as manifestações linguísticas da atualidade se caracterizam por um forte hibridismo, em que fala e escrita, formalidade e informalidade e gêneros textuais se interpenetram numa rede complexa de influências e confluências. Temos de esperar ainda algum tempo para descobrir se o processo de mudança descrito acima — da fala menos monitorada para a escrita mais monitorada — ainda governa as transformações da língua, na era da multimídia em que ler e ouvir, falar e escrever estão se tornando quase a mesma coisa.

De igual modo, a democratização da sociedade brasileira, sua inserção na chamada "globalização", e a recente ascensão social das classes C, D e E — rotuladas de nova classe média — sem dúvida vão alterar os processos de mudança descritos pelos estudos clássicos da sociolinguística. Uma nova classe média traz, em sua ascensão, inovações linguísticas que, se antes eram estigmatizadas, agora já são moeda corrente no PB, sobretudo porque essas camadas sociais representam a maioria da nossa população. O exemplo mais ilustrativo desse fenômeno é a construção para mim + infinitivo, cada vez mais difundida entre falantes das variedades urbanas de prestígio. As relações mais intensas entre as camadas dominantes da sociedade e essa nova classe média provocarão decerto fluxos linguísticos tanto de baixo para cima quanto de cima para baixo.

No que diz respeito também às mudanças implementadas pelas gerações mais jovens, é provável que daqui para a frente as diferenças entre a fala dos mais jovens e dos mais velhos sejam bem menos acentuadas do que antes, graças ao acesso cada vez mais amplo à escolarização e, sobretudo, mais uma vez, aos novos meios de comunicação e às novas tecnologias de informação. A sociedade está ficando cada vez mais informal e, com isso, também a linguagem. A televisão, por exemplo, principal meio de acesso à informação e ao entretenimento da imensa maioria da população brasileira, promove — para o bem e para o mal — um nivelamento da linguagem e uma difusão de novas palavras, novas expressões, novos sentidos e novos modelos sintáticos. Um bordão de programa humorístico, por exemplo, ganha rapidamente seu lugar entre as expressões da língua, o mesmo valendo para modos de falar que caracterizam determinados personagens de telenovelas. Algumas dessas inovações são efêmeras, mas várias se integram permanentemente no repertório linguístico do VGB. Um caso exemplar é o da marca de eletrodomésticos Brastemp que, graças a uma campanha publicitária de vinte anos atrás (década de 1990), se transformou em substantivo comum na locução "não é nenhuma Brastemp".

Além disso [ver capítulo anterior [▸64], os falantes não são "subjugados" pela língua, mas influenciam e interferem diretamente nela. Por exemplo, em inglês, o sintagma *the United States* ('os Estados Unidos') era tratado como plural — *"the United States are a big country"* ('os Estados Unidos são um grande país'). No entanto, depois da Guerra Civil Americana (1861-1865), o verbo no singular — *"the United States is..."* — passou a ser empregado para representar o espírito de união do país e hoje é a forma padronizada, apesar de ir contra as regras da gramática normativa do inglês.

Os meios de comunicação também são um poderoso fator de mudança linguística. Por exemplo: o nome do mafioso italiano Tommaso Buscetta (1928-2000) sempre foi pronunciado na Rede Globo de Televisão como [bus'keta], embora a pronúncia em italiano fosse [bu'ʃeta] — decerto para evitar qualquer semelhança com a palavra-tabu *buceta*. O mesmo vale para a região de Kosovo que, por imposição da mesma rede de televisão, tem sido pronunciado como proparoxítono ['kosovo] e não

como paroxítono, conforme se fazia antes das disputas envolvendo sérvios e albaneses pelo controle da área atraírem o interesse da mídia. Recentemente, no triste 25º aniversário do acidente nuclear de Tchernobyl, na Ucrânia, os locutores da TV Globo passaram a pronunciar [ʧer'nɔbiʊ], como paroxítona, em lugar da pronúncia oxítona que foi usada na época do acidente.

A expressão *risco de vida*, usada há séculos em português, tem sido substituída por *risco de morte* devido à influência de um conhecido comentarista de questões gramaticais cuja projeção midiática o transformou numa (suposta) autoridade sobre assuntos referentes à língua. E a recente polêmica envolvendo o uso do feminino *presidenta*, conforme desejo da própria Dilma Rousseff, é bem reveladora do poder que têm as palavras de revelar as inclinações político-ideológicas dos grupos sociais. Outra polêmica do mesmo tipo ocorreu quando um ministro do governo de Fernando Collor usou o termo *imexível* — perfeitamente possível segundo as regras de derivação da língua, a palavra foi alvo de sarcasmo e condenação; na verdade, um modo de ironizar e discriminar a pessoa que a proferiu. O mesmo se diga da tentativa patética de um ex-governador do Distrito Federal que "demitiu" o gerúndio por decreto.

Tudo o que se faz na área das políticas linguísticas e dos planejamentos linguísticos constitui, de fato, uma intervenção racional, planejada sobre os destinos dos idiomas. A oficialização de uma língua ou dialeto, o resgate de línguas minoritárias por meio de estímulos ao seu uso e difusão, a implantação de alfabetos novos em países onde se escrevia com outros sistemas, a criação de formas escritas para línguas até então ágrafas, a introdução de uma língua no sistema educacional... são todas ações que, consciente ou inconscientemente, atingem não só o *status* sociopolítico como também o próprio sistema lexicogramatical das línguas.

Um caso interessante é o do recente reconhecimento do português uruguaio como uma das línguas que constituem o patrimônio cultural do país. Durante muito tempo perseguido e condenado pelas políticas oficiais — o que fazia seus usuários sentirem vergonha de falá-lo —, o português uruguaio foi promovido em 2008 a língua de ensino. Na região da fronteira Uruguai-Brasil, onde a língua é falada, os estudantes terão aulas de espanhol e de português. Como são muitas as variedades empregadas, o processo de oficialização da língua fatalmente vai resultar na eleição de uma delas para constituir a norma de ensino.

4.6 Conclusão

A mudança linguística, portanto, é fruto da interação permanente e intensa de fatores socioculturais e sociocognitivos. Sua gênese, sua difusão e sua implementação estão estreitamente vinculadas à história social dos falantes de uma língua.

É certo dizer que existem *universais* de mudança linguística. No entanto, como a história de cada comunidade de falantes tem seu ritmo, também as mudanças processadas em sua língua terão ritmos diferentes. É muito interessante examinar variedades de uma mesma língua para detectar esses diferentes ritmos da mudança. No PB, existem variedades que apresentam mudanças mais radicais com relação a outras variedades, mais conservadoras. Enquanto em muitas comunidades a consoante [ʎ], de *palha*, está se deslateralizando e se vocalizando em [ɪ], ['paɪa], essa mesma consoante está ressurgindo em outras comunidades devido a outros fatores de condicionamento linguístico: na variedade falada em Belém do Pará, a consoante [l] quando seguida de [i] tende a adquirir o traço palatal da vogal, de modo que seus falantes dizem [a'ʎi] para *ali*, [ʎi'gar] para *ligar*, ['ʎivru] para *livro* etc. No entanto, nessa mesma variedade (e na maranhense também), a consoante /ɲ/ de *sonho* é pronunciada como [nʲ], de modo que o verbo *sonha* e o nome próprio *Sônia* têm pronúncia semelhante. Com isso, a palavra *linha*, na fala de Belém, soa ['ʎĩnʲa] e não ['lĩɲa] como na maioria das variedades do PB.

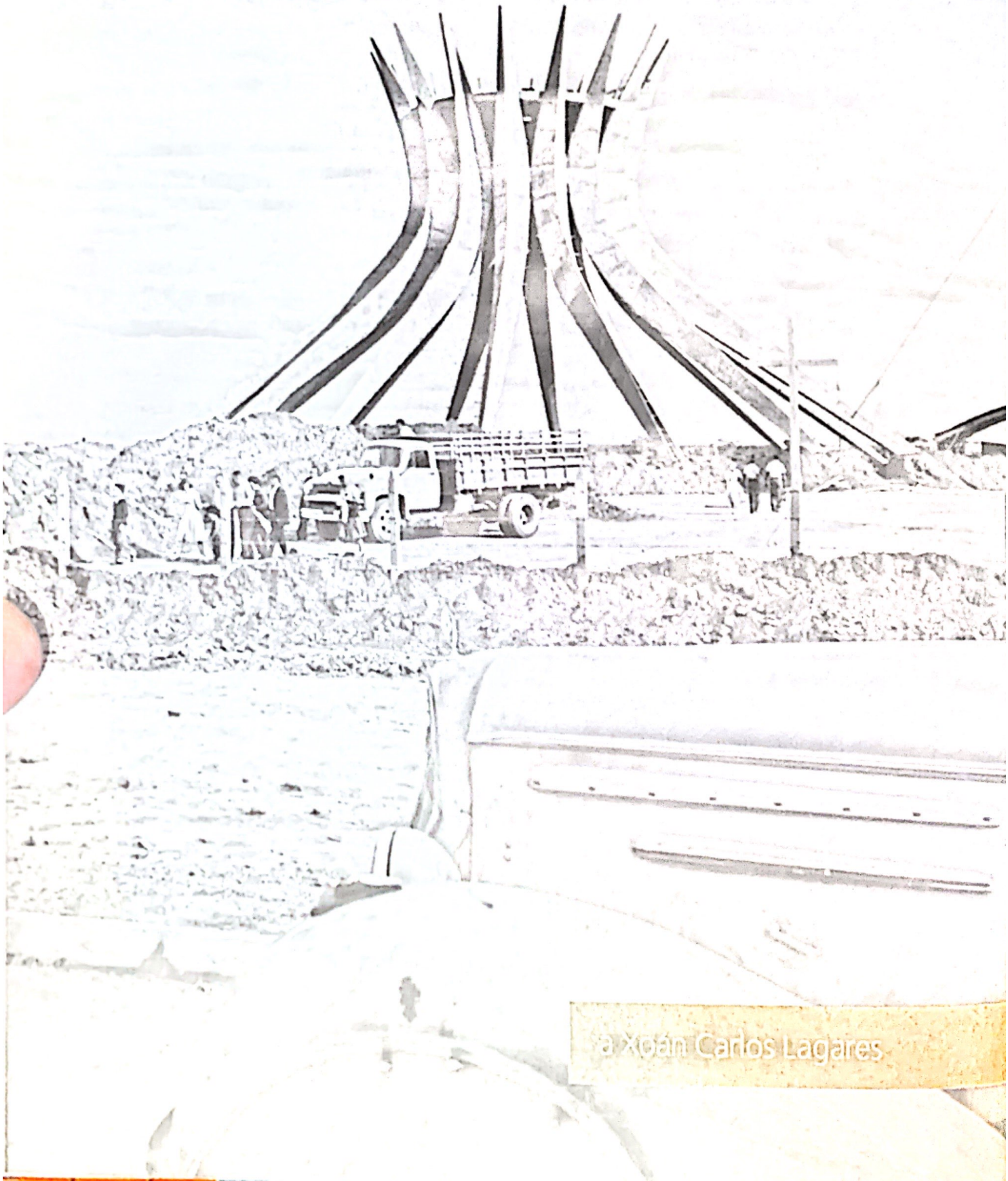
As mudanças linguísticas são inevitáveis, mas podem ser aceleradas, retardadas ou até revertidas, de acordo com fatores socioculturais, sociolinguísticos e sociocognitivos. Mas a reversão é rara, na maioria dos casos a mudança segue em frente.

A ideia de que as línguas "se degeneram" por causa da "corrupção" que os falantes "descuidados" provocam nelas é mais antiga do que andar em pé. O dicionarista inglês Samuel Johnson, autor do famoso *Dictionary of the English Language* (1755), escreveu, no prefácio à sua obra:

> As línguas, como os governos, têm uma tendência natural à degeneração; temos preservado por muito tempo nossa constituição, façamos alguns esforços pela nossa língua.

Essa noção merece ser estudada no que ela tem de cultural, antropológico, ideológico. Mas não pode de modo algum presidir o trabalho daqueles que têm na língua o seu objeto de atuação profissional. No campo da educação, é imprescindível adquirir uma base sólida de informações sobre os processos de mudança linguística para não se deixar levar por um senso comum que, ao fim e ao cabo, só contribui para um sentimento negativo dos falantes com relação à sua língua e, pior, com relação a si mesmos, acusados de deturpadores de um bem precioso.

A língua não existe. O que existe, concretamente, são falantes da língua, seres humanos com história, cultura, crenças, desejo e poder de ação. A língua muda porque os falantes, todos, são dotados de extraordinárias capacidades cognitivas, de um cérebro que o tempo todo, a cada instante, está processando e reprocessando a língua, que é o mais importante vínculo de cada indivíduo com o universo que o rodeia e o mais importante cimento de construção da identidade de um grupo humano.



a Xoán Carlos Lagares

5

do galego ao brasileiro

história da nossa língua

Nesse capítulo, vou usar critérios de classificação das línguas que não correspondem ao que se faz tradicionalmente nos estudos de linguística histórica. Minha classificação se baseia numa concepção de LÍNGUA que não se restringe aos aspectos exclusivamente linguísticos (fonologia, sintaxe, léxico etc.), mas inclui os aspectos históricos, geográficos, culturais, antropológicos, identitários etc. Essa concepção é coerente com a ideia, defendida nessa gramática, de que o PB é uma língua

plena, e não uma "modalidade" ou "variedade" de uma língua chamada genericamente de *português*. Por isso é que, por exemplo, considero também o espanhol argentino, o espanhol peruano, o espanhol mexicano etc. como línguas pertencentes a um grupo chamado *hispânico*.

5.1 O GRUPO PORTUGALEGO NO CONJUNTO DAS LÍNGUAS ROMÂNICAS

O português brasileiro pertence a um grupo de línguas que vamos chamar aqui de portugalego, um nome formado da junção de *português* e *galego*, as duas línguas mais antigas do grupo, embora a ordem cronológica seja inversa à da formação do nome: primeiro nasceu o galego e do galego nasceu o português. Assim, todas as línguas do grupo são continuações históricas do galego falado no noroeste da Península Ibérica, que por sua vez é resultante do contato linguístico do latim vulgar com a(s) língua(s) céltica(s) que eram faladas ali antes da chegada dos romanos e, eventualmente, com outras línguas das quais não nos sobrou nenhuma notícia.

O grupo portugalego, por sua vez, faz parte de uma família de línguas chamada românica, palavra derivada do nome da cidade de Roma, capital do Império Romano, a grande potência militar e política da Antiguidade ocidental. A família românica se compõe de línguas resultantes das transformações sofridas pelo latim nas diferentes regiões conquistadas pelos romanos e anexadas ao Império com o nome de províncias, termo provavelmente derivado do verbo *víncere* ("vencer"): uma província seria uma região *vencida*, isto é, conquistada para (*pro-*) o Império. O latim era a língua dos romanos e seu nome provém de Lácio (em latim: *Latium*), nome da região central da Península Itálica onde se situa a cidade de Roma.

[Mapa 5.1. Línguas faladas na Itália no séc. VIII a.C.]

Além do latim, outras línguas eram faladas na Itália. Algumas eram aparentadas entre si como o osco, o úmbrio, o falisco, o sabélico e o prenestino, que formavam com o latim o grupo *itálico* que, por sua vez, pertence à grande família linguística indo-europeia, que inclui dezenas de línguas faladas na Europa e na Ásia (Irã, Paquistão, Índia).

Também existia o etrusco, a língua de um povo que foi muito importante no norte e

no centro da Itália entre 800 e 500 a.C. Essa língua não era indo-europeia e até hoje desafia os pesquisadores. Os primeiros governantes de Roma foram etruscos e sua língua contribuiu com diversas palavras que até hoje usamos, transmitidas a nós pelos romanos: *abril, vulcão, servo, átrio, calendário, vernáculo, histriônico*.

No sul da Península Itálica, havia colônias gregas (a mais famosa é Nápoles, do grego *nea pólis*, 'cidade nova') que ocupavam a Apúlia e a Calábria (o "salto" e o "bico" da bota italiana, respectivamente) e a Sicília. Essa região era chamada de *Magna Grécia*.

As línguas românicas incluem, além do grupo portugalego que vai nos interessar aqui, outros grupos de línguas: o *itálico*, o *romanche*, o *gálico* (francês, valão, quebequense etc.), o *sardo*, o *corso*, o *occitânico*, o *catalânico*, o *romênico* (romeno, moldavo etc.), o *ásturo-leonês*, o *hispânico* (castelhano, andaluz, ladino, mexicano, cubano, peruano, argentino, paraguaio etc.).

Em todo o mundo, as línguas românicas são faladas, como língua materna e/ou segunda língua, por cerca de 800 milhões de pessoas. Dessas, o grupo hispânico, com suas diversas línguas, é o que tem o maior número de falantes. Se adotarmos o conceito de *língua nacional* como a língua que dá identidade a um povo constituído numa nação independente e soberana, o português brasileiro é a língua românica nacional mais falada no mundo, com mais de 190 milhões de falantes maternos (seguida pelo espanhol mexicano, com cerca de 100 milhões), a segunda língua mais falada no Ocidente (depois do inglês americano, com 300 milhões de falantes) e a quarta mais falada do planeta depois do mandarim (1 bilhão), do híndi (800 milhões) e do inglês americano.

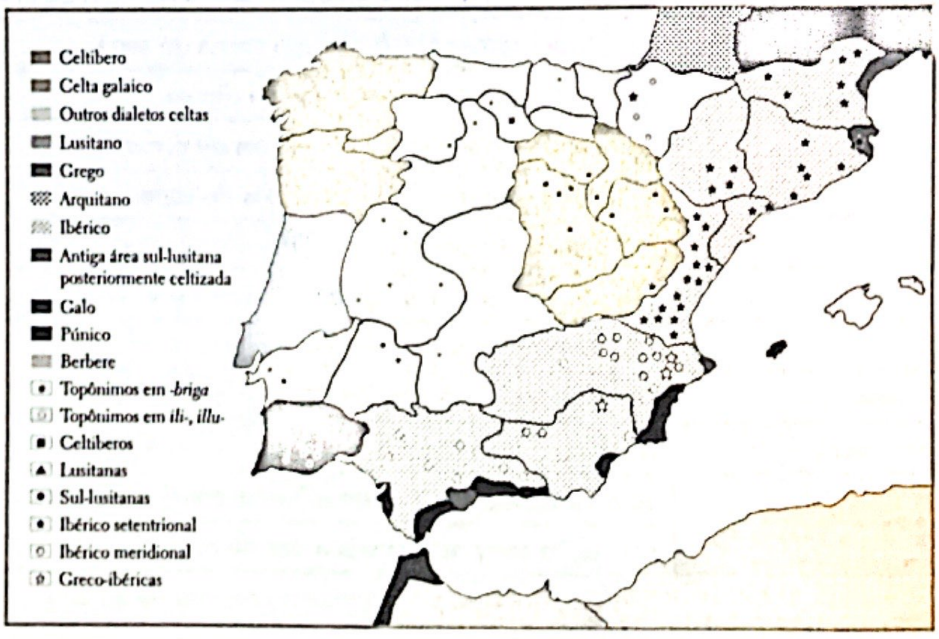
LATIM	*(Illa) claudit semper fenestram antequam cenat.*
ARAGONÊS	*Ella tranca siempre la finestra antes de cenar.*
ASTURIANO	*Ella pieslla siempre la ventana/feniestra primero de cenar.*
BERGAMASCO	*(Lé) la sèra sèmper sõ la finèstra prima de senà.*
BURGUINHÃO	*All farme tôjor lai fenétre aivan de dïgnai.*
CASTELHANO	*Ella siempre cierra la ventana antes de cenar.*
CATALÃO	*Ella tanca sempre la finestra abans de sopar.*
CORSO	*Ella chjudi sempre u purtellu primma di cenà.*
ESTREMENHO	*Ella afecha siempri la ventana endantis e recenal.*
FRANCÊS	*Elle ferme toujours la fenêtre avant de dîner.*
FRANCO-CONDADÊS	*Lèe çhioûe toûedge lai f'nétre d'vaïnt loù dénaie.*
FRANCO-PROVENÇAL	*(Ye) hlou totin a fenetre deant que de cena.*
FRIULANO	*Jê e siare simpri la feneste prime di cenâ.*
GALEGO	*Ela pecha sempre a xanela antes de cear.*
LEONÊS	*Eilla pecha siempres la ventana primeiru de cenare.*

MILANÊS	*(Lee) la sara semper su la finestra primma de disnà.*
MIRANDÊS	*Eilha cerra siempre la bentana/jinela atrás de jantar.*
NAPOLITANO	*Chella sempre chiud' ' a fenesta prima ' e mangià.*
NORMANDO	*A froumae tréjous al crouésie devaunt del dînaer.*
OCCITANO	*Barra totjorn la fenèstra abans de sopar.*
PIEMONTÊS	*Chila a sara sèmper la fnestra dnans ëd fé sin-a.*
PORTUGUÊS	*(Ela) fecha sempre a janela antes de jantar.*
ROMANCHE	*Ella clauda/serra adina la fanestra avant ch'ella tschainia.*
ROMANESCO	*(Quella) chiude sempre 'a finestra prima de magnà*
ROMENO	*Ea închide totdeauna fereastra înainte de cina.*
SARDO	*Issa serrat semper sa bentana antes de chenare.*
SICILIANO	*Idda chiudi sempri ' a finestra avanti ca pistìa.*
TOSCANO	*(Lei) chiude sempre la finestra prima di cenare.*
VALÃO	*Ele sere todi li finiesse divant di soper.*
VENEZIANO	*Ela la sera sempre la fenestra prima de cenar.*

**[Exemplos das diferenças e semelhanças entre algumas línguas românicas
Fonte: http://fr.wikipedia.org/wiki/Langue_romane (acesso em 24/8/2010).]**

5.2 Os romanos na Península Ibérica: latim clássico *vs.* latim vulgar

No século II a. C., as tropas romanas invadiram a Península Ibérica. Lá viviam diversos povos, falantes de diferentes línguas não aparentadas entre si: tartéssios, fenícios, gregos, iberos, celtas, celtiberos. Como sempre faziam depois de conquistar



uma nova província, os romanos enviaram para as terras ibéricas funcionários públicos, magistrados, tropas, colonos, artesãos, agricultores etc., todos responsáveis pela implantação, nos territórios conquistados, da língua dos romanos e também de sua cultura, religião, hábitos, instituições, leis, métodos agrícolas, estilos arquitetônicos etc. Foi assim que o latim, língua de um pequeno povo da região central da Itália, se difundiu por um amplo território várias vezes maior e mais populoso do que o Lácio original. E essa difusão se tornará ainda mais espetacular no futuro, quando espanhóis, portugueses e franceses formarem seus impérios coloniais, onde imporão com maior ou menor sucesso a sua língua românica nacional.

É preciso ter sempre em mente, no entanto, que o latim que era levado pelos colonizadores romanos para as novas províncias do Império não se parecia em nada com o latim que até hoje é estudado nas universidades e nos seminários, usado como língua oficial do Vaticano. Esse *latim escolar e eclesiástico* se baseia única e exclusivamente na *língua literária*, preservada nas obras dos grandes escritores da Roma antiga. Nos estudos linguísticos, ele recebe o nome de latim clássico. O latim clássico era de conhecimento e uso muito restrito: somente as pessoas letradas detinham o domínio dessa língua literária — e se até hoje em muitos lugares do mundo (como o Brasil) o número de pessoas letradas ainda é pequeno com relação ao total da população, na Antiguidade esse número era ainda mais reduzido. Os romanos que se fixavam nas províncias não pertenciam a essa pequena elite letrada, com exceção de alguns poucos altos funcionários. A grande maioria — colonos, soldados etc. — falava diversas variedades regionais que compunham o que costumamos chamar de latim vulgar, isto é, "latim do povo".

As diferenças entre o latim clássico — essencialmente escrito ou falado em ocasiões e lugares marcados por um alto grau de formalidade — e o latim vulgar — essencialmente falado — eram muito grandes, a ponto de constituírem duas *gramáticas* distintas, além de terem todo um *léxico* particular a cada uma delas. Se temos duas *gramáticas* — entendendo gramática como o conjunto das regras que governam o funcionamento da língua — é porque também temos, então, *duas línguas diferentes*.

LATIM CLÁSSICO	LATIM VULGAR	PORTUGUÊS
brassica	caulis	couve
cruor	sanguis	sangue
domus	casa	casa
emere	comparare	comprar
equus	caballus	cavalo
ludere	jocare	jogar
magnus	grandis	grande
os	bucca	boca
pulcher	bellus	belo
sidus	stella	estrela

[Diferenças lexicais entre latim clássico e latim vulgar.]

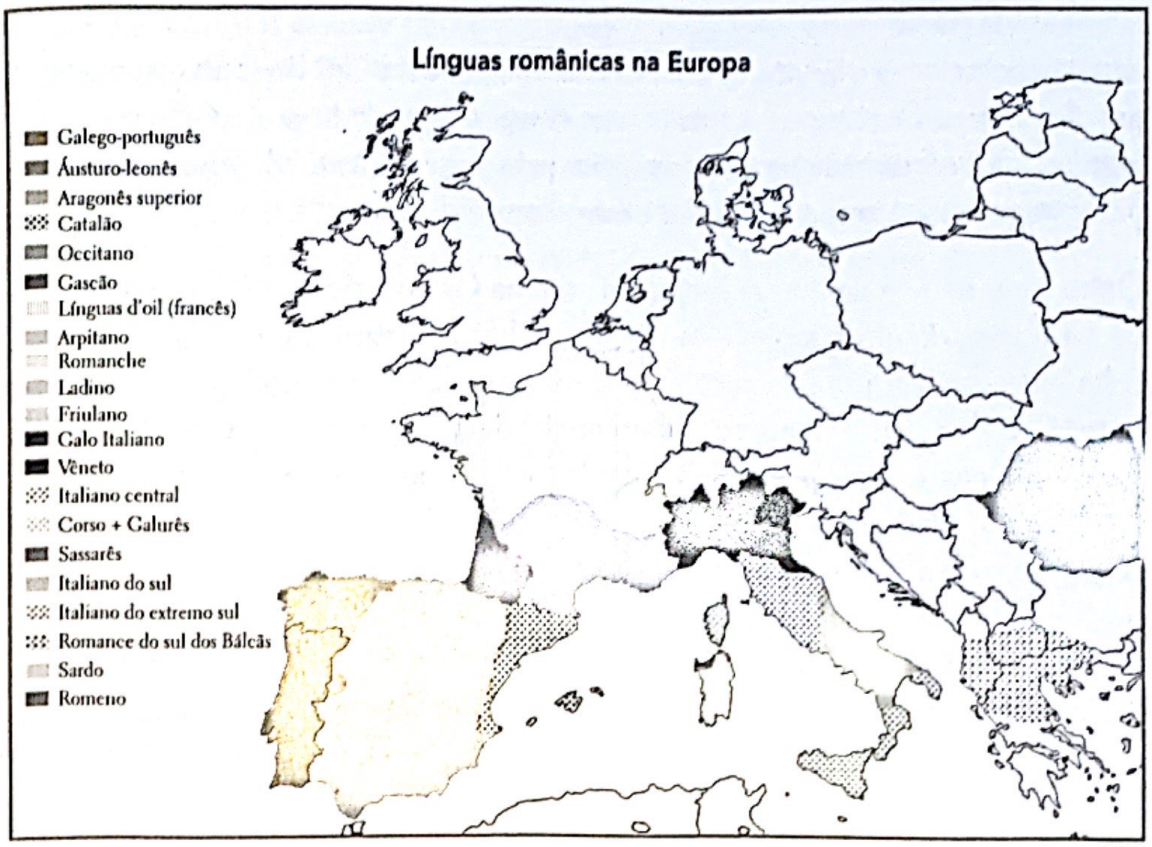
Essas diferenças entre o latim clássico e o latim vulgar são da mesma natureza das diferenças entre o modelo de "língua portuguesa" descrito e prescrito pelas gramáticas normativas (modelo elaborado sempre com base na tradição literária) e o português brasileiro contemporâneo, vivo, mesmo em suas VUP escritas. Basta lembrar que essas gramáticas ainda apresentam a conjugação verbal clássica, como deveria ser o português europeu no século XV, com o índice de pessoa *vós* (que deixou de ser usado há cerca de trezentos anos), desconsiderando totalmente as grandes mudanças ocorridas no quadro dos índices de pessoa e não-pessoa e, por conseguinte, na morfologia verbal, não só no português brasileiro como também no português europeu e nas outras línguas do grupo.

5.3 De novo, o contato linguístico

Os povos conquistados pelos romanos na Europa ocidental acabaram por adotar o latim vulgar e por abandonar, no espaço de algumas gerações, suas línguas ancestrais. Como sempre ocorre quando uma população adota uma nova língua, essa adoção provoca mudanças na língua adotada: o povo que passa a falar a língua nova transfere aspectos linguísticos de seu idioma original para a língua nova — hábitos articulatórios (os modos de pronunciar as palavras), regras morfossintáticas (a gramática propriamente dita), léxico (palavras). Essas mudanças são resultantes, portanto, do contato linguístico [▶133] e explicam, por exemplo, por que as línguas românicas são ao mesmo tempo parecidas entre si, mas também diferentes: basta comparar as línguas ibéricas como o galego, o leonês, o português europeu, o castelhano, o andaluz etc. para verificar isso.

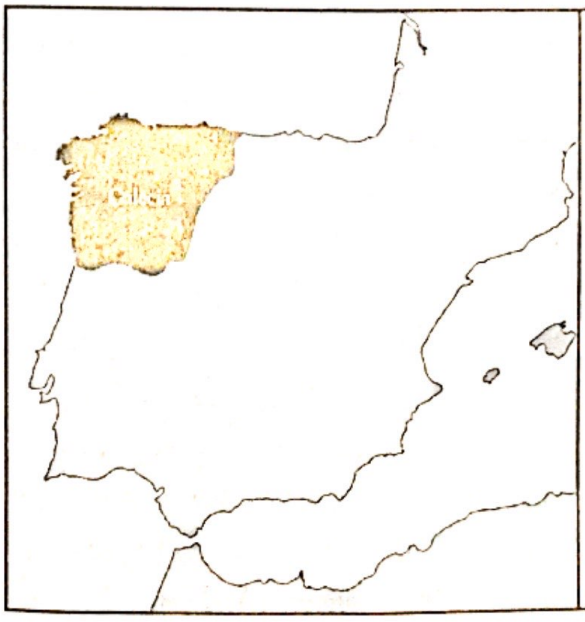
As línguas originalmente faladas nos territórios incorporados ao Império — línguas que, no caso das conquistas romanas na Europa ocidental, desapareceram — formam o que se chama de substrato. A língua que passa a ser falada e se torna hegemônica no conjunto da população compõe o estrato, e é ele que vai determinar a classificação de uma língua dentro de uma família. Dizemos que o galego é uma língua românica porque seu estrato é o latim vulgar.

Em termos botânicos, o substrato é o solo onde uma planta se enraíza para depois crescer para o alto. Assim como as raízes da planta ficam ocultas aos nossos olhos, da mesma forma o substrato linguístico, depois que a língua é abandonada por seus falantes, fica "enraizado" na língua nova, sem que possamos ter acesso direto a ele. Por isso, as influências do substrato na língua só podem ser deduzidas, hipotetizadas. São esses diversos substratos que explicam as diferenças entre as línguas românicas, mesmo entre línguas tão próximas como, por exemplo, o galego e o castelhano, o provençal e o catalão, o toscano e o lombardo etc.

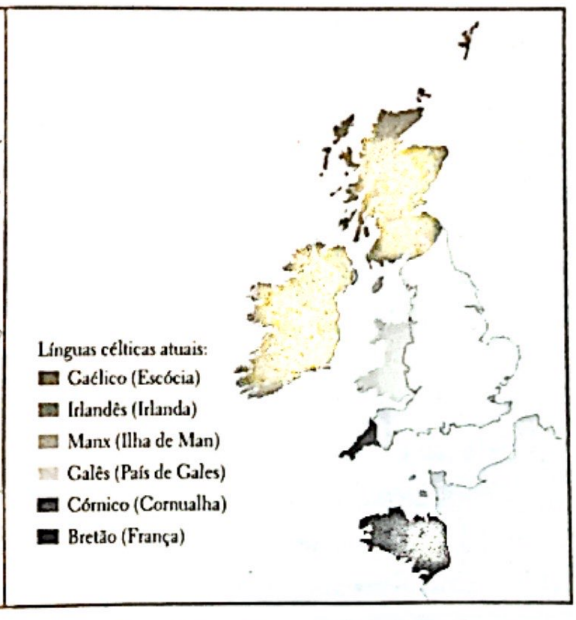


[Mapa 5.3 — Línguas românicas na Europa.]

No extremo noroeste da Península Ibérica, ficava uma região que os romanos passaram a chamar de Gallaecia, designação derivada do nome do povo que ali vivia, os *gallaeci*, que os gregos chamavam de *kallaikoi*. Não sabemos precisamente quem eram esses galaicos que habitavam aquela região nem como era a língua (ou as línguas) que falavam. Tudo parece indicar que era uma língua da família céltica, uma família de línguas que foram muito importantes na Antiguidade, mas que aos poucos foram desaparecendo — suas descendentes atuais sobrevivem como línguas



[Mapa 5.4 — A província romana da Galécia.]



[Mapa 5.4 — Línguas célticas atuais.]

minoritárias em alguns pontos da Europa: o bretão, na França; o galês e o gaélico, na Grã-Bretanha; o irlandês na Irlanda. Na Irlanda e na Grã-Bretanha, sob a forte pressão do inglês, as línguas célticas quase desapareceram e hoje são faladas por um grupo muito reduzido de pessoas que lutam pela preservação dos idiomas ancestrais. Na França, o bretão também sofreu e sofre a pressão do francês.

Outro forte indício de que as línguas faladas na Galécia eram célticas é a presença, na região, de vários topônimos (nomes de lugar) terminados em *-briga*, palavra celta que significava 'fortificação', como é o caso de *Conímbriga*, que resultou no nome *Coimbra*. São também célticos os nomes dos rios portuguesess Douro (de *dur*, 'água'), Minho e Tâmega. Esse último pode ter relação com o nome do rio Tâmisa, que banha a cidade de Londres, com a provável significação de 'escuro' (há também os rios Támoga e Tambre, na Galiza, de mesma origem). Outra evidência céltica em topônimos é o nome da cidade de Évora (*Ébora*, em latim). Esse nome pode derivar do teixo, árvore importante para os celtas, que a usavam para envenenar suas flechas. O nome do teixo em irlandês moderno é *ibhar*. A cidade inglesa de York era chamada em latim de *Eburacum* e tem a mesma origem, assim como as francesas Ébreuil (*Eboralacensis*) e Évreux (*Eburovives*). Por fim, o próprio nome Portugal parece ser a junção do latim *Portus* ('porto') e do celta *Cale*, nome de uma povoação que se localizava onde hoje existe a cidade de Vila Nova de Gaia (Gaia provém de *Cale*), na margem oposta do rio Douro à cidade do Porto.

O povo daquela região, os galaicos, adotou o latim vulgar. Imaginamos que sua língua fosse da família céltica por causa de determinadas características que existem até hoje nas línguas dessa família e que se manifestaram nas línguas românicas que passaram a ser faladas nas regiões anteriormente habitadas por povos celtas (a França atual, a Península Ibérica — Espanha e Portugal —, o norte da Itália). É o caso da lenização das consoantes intervocálicas, conforme já vimos [▶137].

Esse tipo de mudança linguística resulta, portanto, do contato de línguas: povos de línguas diferentes entram em interação, se influenciando mutuamente [▶133]. Quando um povo inteiro adota a língua do outro, é comum haver transferências de hábitos linguísticos da língua original, mesmo que ela desapareça. Essas questões de contato linguístico serão muito importantes mais adiante, quando a gente for analisar as mudanças ocorridas no PB e que transformaram ele numa língua diferente do português europeu, da qual o PB deriva historicamente.

O PB é o que é — uma língua distinta das demais do grupo portugalego — porque muitas de suas características parecem decorrer do uso e da transformação do português durante o longo período colonial pela grande população de escravos africanos arrancados de suas terras de origem e forçados a vir para cá. Desse contato do português com as línguas africanas, principalmente as do grupo banto, resultou em grandes traços o PB, especialmente (mas não exclusivamente) em suas variedades "populares", isto é,

as empregadas pela imensa maioria da nossa população. Os grandes deslocamentos de escravos pelo território brasileiro durante os ciclos econômicos do período colonial (das plantações de cana de açúcar no Nordeste para a extração do ouro nas Minas Gerais, por exemplo) explicam por que o PB falado pelas camadas populares das mais diferentes regiões apresenta uma grande uniformidade em seus principais aspectos morfossintáticos, com destaque para as regras de concordância nominal e verbal.

5.4 As línguas do grupo portugalego

Os *gallaeci* legaram seu nome ao povo que tem vivido na Galiza, os *galegos*. É na Galiza que vai surgir a língua da qual vão se originar todas as línguas que hoje incluímos no grupo *portugalego*, entre elas a nossa, o português brasileiro.

O grupo portugalego compreende as seguintes línguas:

LÍNGUA	LOCALIZAÇÃO PRINCIPAL
angolar	São Tomé (ilha de)
cabo-verdiano	Cabo Verde (África)
crioulo da costa de Coromandel	Coromandel (Índia)
crioulo de Ambom	Ambom (Indonésia)
crioulo de Bengala	Bengala (Índia)
crioulo de Bombaim	Bombaim (Índia)
crioulo de Flores	Flores (Indonésia)
crioulo de Java	Java (Indonésia)
crioulo de Korlai	Korlai (Índia)
crioulo de Macassar	Macassar (Indonésia)
crioulo de Macau (macaense)	Macau (China)
crioulo de Quilom	Quilom (Índia)
crioulo de Tellicherry	Tellicherry (Índia)
crioulo de Ternate	Ternate (Indonésia)
crioulo de Trincomalee e Batticaloa	Trincomalee e Batticaloa (Sri Lanka)
cristang	Malásia (Ásia)
fá d'Ambô	Ano Bom (África)
forro (santomense)	São Tomé (África)
galego	Galiza (Espanha)
kriol	Guiné-Bissau (África)
língua da casa	Damão (Índia)

língua dos velhos	Diu (Índia)
lunguyê	Príncipe (África)
papiamento	Curaçao, Aruba, Bonaire (Antilhas)
patuá macaense	Macau (China)
português angolano	Angola (África)
português brasileiro	Brasil (América do Sul)
português cabo-verdiano	Cabo Verde (África)
português europeu	Portugal (Europa)
português guineense	Guiné-Bissau (África)
português moçambicano	Moçambique (África)
português santomense	São Tomé e Príncipe (África)
português timorense	Timor Leste (Indonésia)
português uruguaio	Uruguai (América do Sul)

A situação sociopolítica de cada uma dessas línguas apresenta suas peculiaridades. A condição de língua materna hegemônica, além de língua oficial do Estado, só existe no português brasileiro e no português europeu.

Essa condição, no entanto, não deve ser confundida com monolinguismo, uma ideologia muito presente na sociedade brasileira, onde historicamente se tem ocultado a realidade plurilíngue do nosso país. Além do PB, certamente hegemônico, são faladas cerca de 200 outras línguas, a maioria delas indígenas (180 aproximadamente), junto com línguas trazidas por imigrantes europeus e asiáticos (alemães, italianos, japoneses, coreanos, sírio-libaneses, poloneses, ucranianos, espanhóis, galegos etc.), além dos resquícios de línguas africanas. Na fronteira do Brasil com o Paraguai, ocorre uma penetração do espanhol paraguaio e do guarani em terras brasileiras, bem como um uso cada vez mais intenso do PB em território paraguaio. No extremo norte do Brasil, no Amapá, na região limítrofe com a Guiana Francesa, se usa uma língua crioula de base francesa denominada *lanc patoá*. Uma intensa imigração de bolivianos para o Brasil, sobretudo para a cidade de São Paulo, faz crescer o uso do espanhol boliviano, do quéchua e do aimará naquela metrópole.

O português uruguaio é empregado no norte do Uruguai, nas áreas limítrofes com o estado do Rio Grande do Sul. Não tem origem em algum tipo de migração de brasileiros para aquele país: ao contrário, é uma língua ancestral, falada há séculos naquela área, que foi incorporada ao território uruguaio no processo de independência do país. Durante muito tempo, foi alvo de perseguição, condenação e combate por parte da administração política e educacional uruguaia, de modo que seus falantes sempre se sentiram desprestigiados e marginalizados. Recentemente, porém, uma nova lei nacional de educação (2008) reconheceu o caráter plurilíngue da nação uruguaia e lançou dispositivos legais para o ensino *de* e *em* português uruguaio na região fronteiriça.

Em Portugal, a língua mirandesa, do grupo ásturo-leonês, foi oficializada em 1999, e conta com cerca de 13 mil falantes na região de Miranda do Douro. Abaixo, uma amostra dessa língua (publicada no jornal português *Público*, em 24 de julho de 2007):

Muitas lhénguas ténen proua de ls sous pergaminos antigos, de la lhiteratura screbida hai cientos d'anhos i de scritores hai muito afamados, hoije bandeiras dessas lhénguas. Mas outras hai que nun puoden tener proua de nada desso, cumo ye l causo de la lhéngua mirandesa.	Muitas línguas têm orgulho dos seus pergaminhos antigos, da literatura escrita há centenas de anos e de escritores muito famosos, hoje bandeiras dessas línguas. Mas há outras que não podem ter orgulho de nada disso, como é o caso da língua mirandesa.

O galego é falado por cerca de 3 milhões de pessoas na região autônoma da Galiza, no noroeste da Espanha. Graças à redemocratização da Espanha, a partir de 1976, após quatro décadas de ditadura franquista, o galego se tornou língua oficial da região autônoma, ao lado do castelhano.

Nas ex-colônias portuguesas da África, o português sempre enfrentou a concorrência de outras línguas, mais amplamente usadas pela população. Em alguns países, seu uso é praticamente restrito à administração pública e ao ensino, o que explica por que, aos ouvidos brasileiros, as formas que o português assumiu naquele continente soem parecidas com o português europeu, que serve de base para o padrão a ser usado pelos órgãos oficiais e pelo sistema educacional.

Em Angola, cerca de 80% da população se considera falante do português angolano: nesse país são faladas várias línguas, das quais as mais difundidas são o quimbundo, o umbundo e o quicongo, todas da família banta e que tiveram papel importante na formação histórica do PB. Na Guiné-Bissau, somente 14% da população domina o português guineense: a grande maioria da população fala o kriol (ou crioulo da Guiné-Bissau), língua que também se emprega no vizinho Senegal, na região de Casamansa. O kriol é reconhecido como língua nacional, mas não goza de estatuto oficial.

No arquipélago de Cabo Verde, o português cabo-verdiano se restringe aos usos administrativos: a grande maioria da população emprega o kabuverdianu, que também é falado pela diáspora cabo-verdiana mundo afora (mais de 400 mil emigrados, dos quais 250 mil só nos Estados Unidos, enquanto a população residente em Cabo Verde soma pouco mais de 500 mil pessoas). O kabuverdianu é bastante diversificado, com variedades próprias a cada uma das nove ilhas habitadas (de um total de dez). Não dispõe de estatuto oficial, embora haja negociações para que seja co-oficializado ao lado do português.

Na república formada pelas ilhas de São Tomé e Príncipe, o português santomense é a língua oficial do Estado, empregada por grande parte da população. Ao lado dele convivem outras três línguas: o forro (na maior parte da ilha de São Tomé), o angolar

(falado por uma pequena comunidade no extremo sul de São Tomé) e o lunguyê (falado na ilha do Príncipe, razão por que também é chamado de principense).

No mesmo Golfo da Guiné onde se localiza São Tomé e Príncipe, fica a ilha de Ano Bom, que foi possessão portuguesa durante muito tempo, antes de passar ao domínio espanhol no final do século XVI. Hoje pertence à República da Guiné Equatorial. Nessa ilha se fala o fá d'ambô ("fala de Ano Bom"), que se assemelha em grande medida ao forro de São Tomé, embora já com forte influência do espanhol e do inglês.

Em Moçambique, apenas 6% da população se diz falante do português moçambicano como língua materna. Os demais habitantes do país empregam uma ou várias das mais de quarenta línguas do grupo banto existentes no território.

No Timor Leste, somente 5% da população usa o português timorense, sendo essa população essencialmente idosa. O desaparecimento do português em Timor Leste se deve à invasão que a Indonésia promoveu no país quando ele se declarou independente de Portugal em 1976. O exército indonésio se apoderou de Timor Leste e promoveu um verdadeiro genocídio, assassinando cerca de ¼ da população total. O português foi proibido e reprimido, junto com a religião católica. Somente 25 anos depois, graças a fortes pressões internacionais pela libertação do país, foi proclamada a República Democrática do Timor Leste. A língua mais difundida é o tétum (da família austronésia). Além dela, outras quinze línguas são empregadas no território, que tem cerca de 800 mil habitantes.

Todas as demais línguas do grupo faladas na Ásia são extremamente minoritárias, algumas já em franca via de extinção, como as de Diu e Damão, na Índia, conservadas apenas pelos mais idosos, sem que tenham sido transmitidas às gerações mais novas. O papiá kristang (de "papear", isto é, "falar", e "cristão") tem cerca de 3 mil falantes na península malaia. Na antiga colônia portuguesa de Macau, devolvida à China em 1999, o português sempre foi língua minoritária e hoje é falado por cerca de 0,2% de uma população de 540 mil habitantes. O crioulo macaense está praticamente extinto.

O papiamento é a língua mais empregada nas Antilhas Holandesas: Aruba, Bonaire e Curaçao. Embora ali o holandês seja a língua oficial mais antiga, a maioria da população emprega com muito mais intensidade o papiamento, que, em 2007 se tornou língua oficial das Antilhas Holandesas. O nome da língua deriva do verbo "papear", isto é, "falar". O papiamento tem cerca de 300 mil falantes. Sua origem se vincula aos escravos trazidos pelos holandeses para a cultura da cana-de-açúcar em suas possessões nas Antilhas. Esses escravos provinham da Guiné-Bissau, de Cabo Verde e de São Tomé e já eram falantes de suas línguas derivadas do português. A partir do século XIX, o papiamento vai receber muita influência do espanhol venezuelano, devido à proximidade das ilhas à costa da Venezuela, e do inglês falado nas ilhas vizinhas de Trinidad e Tobago.

5.4.1 Amostras de algumas línguas do grupo

Assim como existem línguas românicas bastante próximas — como o português brasileiro e o castelhano — e muito diferentes entre si (veja abaixo a oração do pai-nosso em latim, romeno e romanche, língua falada na Suíça) —, no grupo portugalego acontece a mesma coisa. O galego e o português brasileiro, por exemplo, são bastante próximos do ponto de vista fonético, e falantes dessas duas línguas conseguem manter um bom nível de intercompreensão.

LATIM

1. Pater Noster, qui es in caelis,
2. sanctificetur nomen tuum.
3. Adveniat regnum tuum.
4. Fiat voluntas tua, sicut in caelo et in terra.
5. Panem nostrum quotidianum da nobis hodie,
6. et dimitte nobis debita nostra
7. sicut et nos dirnittimus debitoribus nostris.
8. Et ne nos inducas in tentationem,
9. sed libera nos a malo.
10. Amen.

ROMENO

1. Tatăl nostru care ești în ceruri,
2. sfințească-se numele Tău.
3. Vie împărăția Ta,
4. Facă-se voia Ta, precum în cer așa și pe Pământ.
5. Pâinea noastră cea de toate zilele, dă-ne-o nouă astăzi
6. și ne iartă nouă greșelile noastre
7. precum iertăm și noi greșiților noștri,
8. și nu ne duce pe noi în ispită
9. ci ne izbăvește de cel rău.
10. Amin.

ROMANCHE

1. Bab noss, ti che es en tschiel,
2. sanctifitgà vegnia tes num,
3. tes reginavel vegnia tar nus.
4. Tia veglia daventia sin terra sco en tschiel.
6. Noss paun da mintgadi dà a nus oz.
6. Ed ans perduna noss debits,
7. sco era nus perdunain a noss debiturs.
8. E n'ans maina betg en empruvament,
9. ma spendra nu dal mal.
10. Amen.

O kabuverdianu, por seu turno, é uma língua muito diferente do português europeu e do português brasileiro e incompreensível para os falantes dessas línguas. Veja um exemplo do kabuverdianu, a letra de uma *morna*, gênero musical típico de Cabo Verde, interpretada pela cantora Cesaria Evora, responsável pela difusão da língua internacionalmente:

Bô ca tá pensá, nha cretcheu
Nem bô ca ta imaginá
C'ma longe di bô m'tem sofrido
Pregunta lua na céu, lua nha companheira di solidão
Lua, vagabundo di espaço, qui conchê tudo nha vida
Nha desventura,
El qui ta conta-bo, nha cretcheu
Tudo que'm tem sofrido
na ausência i na distância [...]

(letra e música de B. Leza)

Você não pensa, meu querido,
nem imagina
o que longe de você eu tenho sofrido.
Pergunte à luà no céu, à lua, minha companheira de solidão.
Lua, vagabunda do espaço, que conhece toda a minha vida,
a minha desventura,
ela que te conta, meu querido,
tudo o que eu tenho sofrido
na ausência e na distância [...]

A seguir, a oração do Pai Nosso em cristang:

1. Pai nussa ki teng na seu
2. Santafikadu bossa Nomi
3. Bossa reinu beng
4. Bossa bontadi kumpri na terea assi kuma na seu
5. Dar nussa pang di kada dia
6. Perdua nussa pekadu
7. Assi nus perdua nussa inimigu
8. Nag dessa nus teng mal itentasang
9. Mass libra kum nus di tudu mal
10. Amen.

Por fim, o preâmbulo da Declaração dos Direitos do Homem em forro:

> Ca ploclamá Declaraçón Universal di Dirêtu di Hómé sé cuma ua cuá buá pa tudu pôvô cu téla di mundu pô fé, pa tudu nguê cu tudu organizaçón di téla, pô na txilé ni cabeça fa, fé tudu cuá cu nen cá pôji, pá melhorá ensino cu educaçón, disenvolvê lispêtu di dirêtu cu liberdade, tlabá pa pô concê e pa pô platicá tudu cuá sé ni tudu Estado memblu e ni téla cu naxi tê dependenxa fa.

Como se vê, é mais do que justificado considerar a existência de uma família de línguas derivadas do galego e, mais tarde, do português europeu medieval tardio, da mesma forma como os tratados de linguística histórica consideram a existência de uma família de línguas indo-europeias e, a partir de cada uma dessas línguas, a existência de subfamílias de línguas como a românica, a germânica, a eslava, a céltica, a báltica etc. Passado meio milênio da expansão marítima de Portugal, é forçoso reconhecer esse processo de diversificação da língua implantada em outros territórios fora da Europa, por força da mudança linguística que, sendo da própria natureza sociocognitiva da linguagem, é por isso mesmo, inevitável.

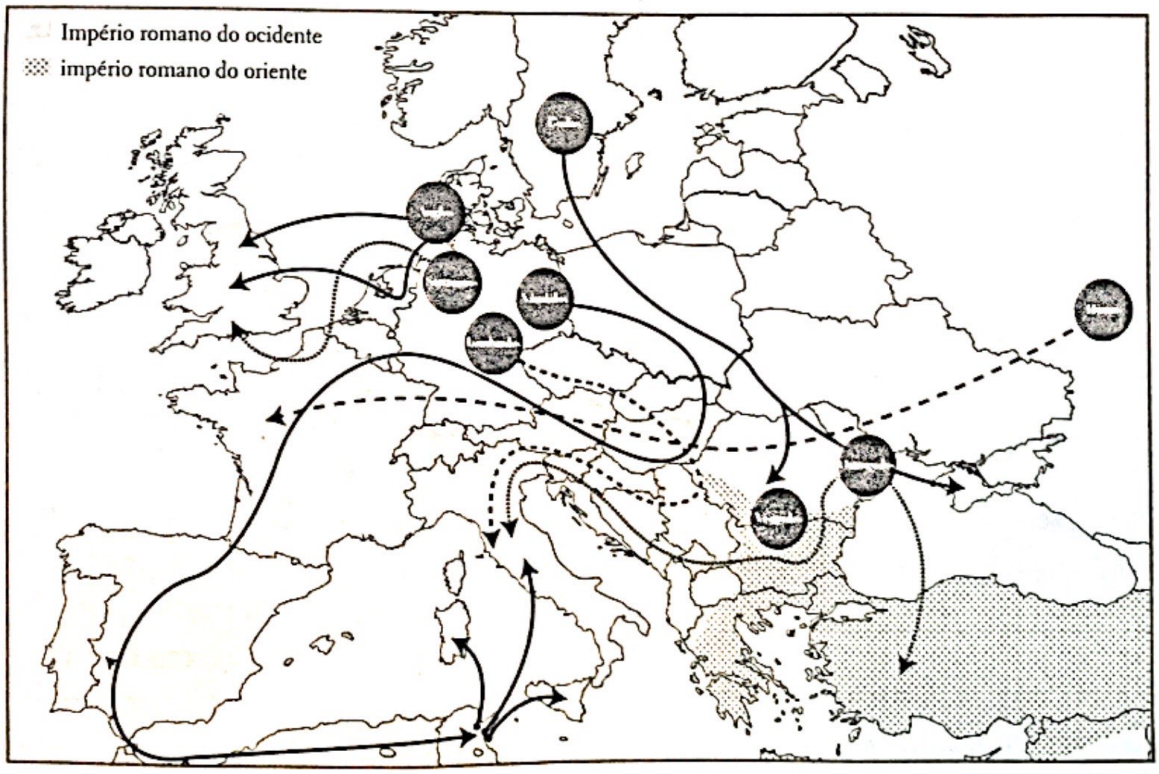
5.5 História linguística da Península Ibérica

O galego surgiu, portanto, no extremo noroeste da Península Ibérica, resultante da evolução do latim vulgar em contato com a(s) língua(s) falada(s) pelos habitantes originais da região. Mais tarde, a partir do século V da nossa era, a Península Ibérica foi invadida por povos de línguas germânicas (vândalos, suevos e visigodos) que romperam a unidade política do Império Romano e formaram reinos independentes. Os suevos se fixaram principalmente na Galiza e no norte de Portugal.

No entanto, ao contrário dos romanos, que conseguiram impor sua língua aos habitantes das províncias, esses conquistadores germânicos em poucas gerações adotaram a língua românica falada pelas populações de seus reinos (junto com o cristianismo que, alguns séculos antes, tinha se transformado na religião oficial do Império Romano). As diferentes línguas germânicas constituíram o que se chama de superestrato: uma lín-

gua de um povo conquistador que não se impõe ao povo conquistado e que não influi decisivamente nos destinos da língua dos conquistados, deixando nela principalmente contribuições lexicais. Assim, são de origem germânica palavras como *arauto, arreio, agasalho, albergue, anca, aspa, barão, banco, bando, bandeira, banho, branco, brasa, brandir, dardo, esgrimir, espiar, elmo, espeto, estaca, estribo, espora, estampar, escarnecer, feudo, fato, feltro, fralda, fresco, ganso, garbo, guarda, guerra, grupo, galardão, guia, lata, lasca, liso, marco, morno, rico, roupa, roubar, saga, saque, sopa, tirar, trepar, trégua* e os pontos cardeais *norte, sul, leste, oeste*. Também são de origem germânica muitos nomes próprios que se firmarão nas línguas ibéricas, como os seguintes:

ELEMENTO DE COMPOSIÇÃO	ORIGEM E SIGNIFICADO	EXEMPLOS
-ardo	*hart*: "duro", "forte"	Ricardo, Gerardo, Bernardo
-berto, Bert-	*bert*: "brilhante"	Adalberto, Alberto, Roberto, Berta
-fredo, Fred-	*fried*: "paz"	Frederico, Godofredo, Valfredo, Manfredo
-gildo	*gild*: "valor"	Gildo, Astrogildo, Hermenegildo
-ildo, Hild-	*hild*: "combate"	Hildo, Hildebrando, Hildefonso, Romildo
-mundo	*mund*: "proteção"	Edmundo, Raimundo, Sigismundo
-rico, -rigo, -rique, Ric-	*rik*: "rico", "poderoso"	Ricardo, Frederico, Américo, Rodrigo, Henrique, Manrique
-ulfo, -olfo, Alf-	*wulf*: "lobo"	Alfonso, Rodolfo, Ataulfo, Ranulfo, Adolfo, Astolfo



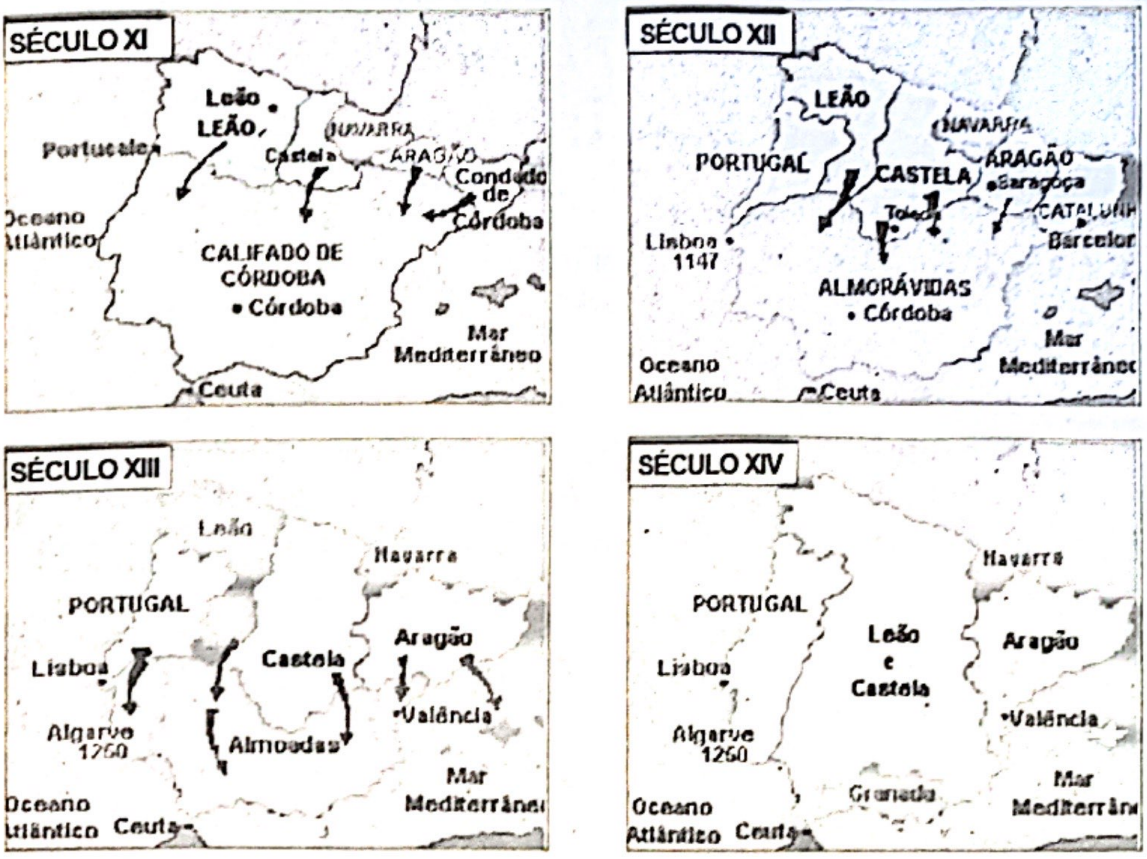
[Mapa 5.5 – Movimentos migratórios e conquistas dos povos germânicos (os hunos, vindos da Ásia, não são germânicos).]

Também de origem germânica é a terminação *-es* (do germânico *-ici*) que se apunha aos nomes próprios masculinos para indicar filiação. Essa é a origem dos conhecidos sobrenomes *Fernandes, Gomes, Gonçalves, Rodrigues, Álvares, Lopes, Marques, Soares, Mendes, Peres* etc., que têm suas formas correspondentes em castelhano: *Fernández, Gómez, González, Rodríguez, Álvarez, López, Márquez, Suárez, Menéndez, Pérez* etc. Por isso é que o primeiro rei de Portugal se chamava *Afonso Henriques*, isto é, Afonso, filho de Henrique.

É preciso lembrar que na Idade Média não existiam nomes de família, as pessoas tinham apenas seu nome de batismo e somente os grandes senhores feudais ostentavam uma designação além do prenome, geralmente extraída de seu local de origem ou de sua posse. O pai de Afonso Henriques, por exemplo, era chamado Henrique de Borgonha, porque era um nobre vindo dessa região da França. Assim, tal como nas comunidades do interior do Brasil é comum designar alguém como "o Severino de Maria", "a Rita de Francisco" etc., aludindo à mãe ou ao pai da pessoa, também em várias regiões da Europa a primeira forma de identificar uma pessoa era mencionando o nome de seu pai. Essa é a origem, por exemplo, dos nomes em Mac (MacDonald, MacMillan, MacMurray etc.), na Escócia (*mac* é 'filho' em gaélico) e em O' (O'Hara, O' Connor, O'Donald etc.), na Irlanda (do irlandês *ó*, originalmente 'neto'). Na Islândia até hoje as pessoas não têm nome de família, mas apenas o nome de batismo seguido do nome do pai com as terminações *-sonn* ('filho') ou *-dóttir* ('filha'), como na Idade Média (a primeira-ministra da Islândia em 2010 se chamava *Jóhanna Sigurðardóttir*, Joana, filha de Sigurd). É assim que sabemos, por exemplo, que o pai de Pedro Álvares Cabral se chamava Álvaro. Com o passar do tempo e com o desenvolvimento do capitalismo, que precisava identificar seus consumidores, e dos direitos de herança, que se estenderam aos burgueses, os antigos patronímicos se transformaram em sobrenomes de família e passaram a ser transmitidos de geração em geração, junto com os bens da família.

Uma consequência do contato das línguas germânicas com os romances falados no centro da Península Ibérica foi a ditongação das vogais abertas /ɛ/ e /ɔ/, quando tônicas, fenômeno que resultou nas formas tão características do castelhano como *puerta, piedra, suerte, siete, nueve, fuerte, cierto* (*porta, pedra, sorte, sete, nove, forte, certo*), enquanto no galego as vogais abertas tônicas se mantiveram. O mesmo fenômeno se observa em francês: *pierre* ('pedra'), *feu* ('fogo'), *neuf* ('nove'), *oeuf* ('ovo'), *loup* ('lobo') etc. — embora esses ditongos tenham voltado a se monotongar no francês moderno (com exceção de *-ie-*), a ortografia conservadora nos mostra que eles existiram durante muito tempo na língua falada. Também no toscano encontramos *puorta, buono, nuovo, pietra, fuoco* etc. Na verdade, o galego (e o português vindo dele) é a única língua românica que conserva as vogais abertas tônicas do latim vulgar, que nas outras línguas se ditongaram.

No ano de 711, ocorre nova invasão. Dessa vez são os mouros islamizados que, vindos do norte da África, ocupam a Península Ibérica quase inteira, com exceção do extremo norte-noroeste, que será o foco de resistência das populações cristãs ao domínio muçulmano e de onde partirá o movimento denominado Reconquista.



[Mapa 5.6 – Reconquista da Península Ibérica pelos cristãos (711).]

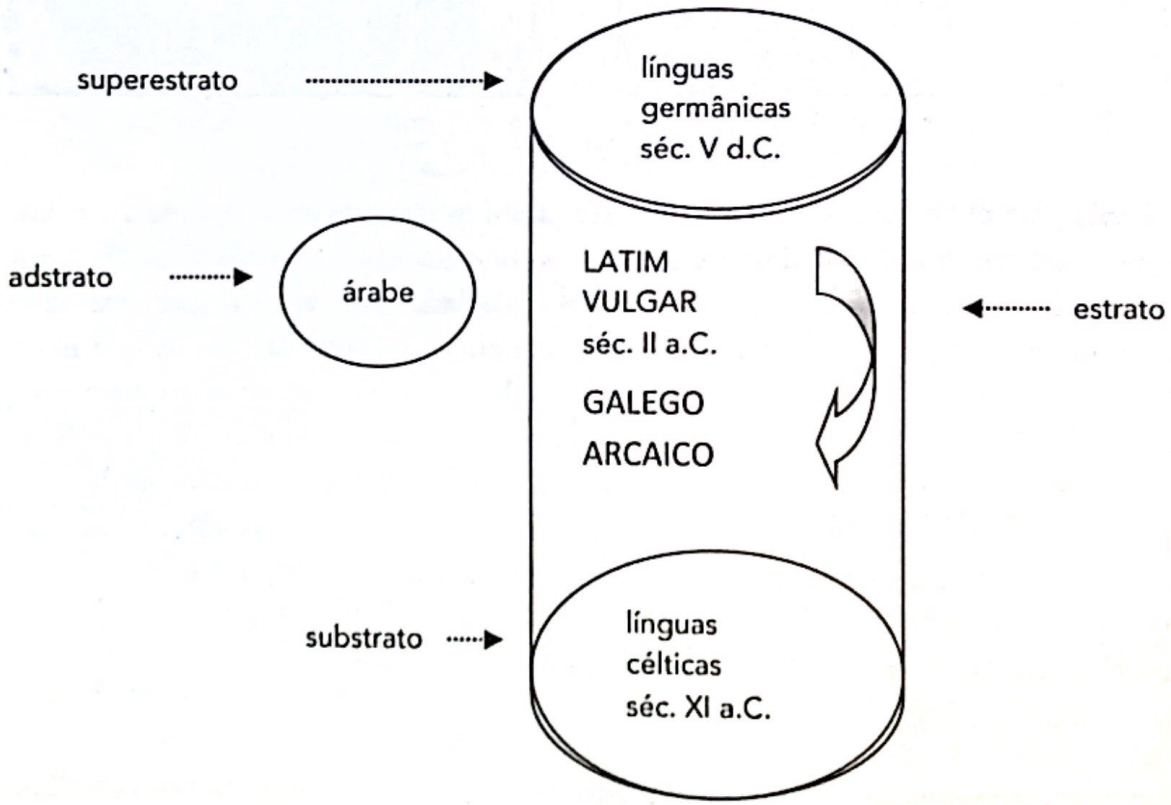
Com a língua árabe agora a situação será outra: os mouros não impõem a sua língua, mas também não assumem a língua dos conquistados — o árabe e as diferentes variedades românicas ibéricas conviverão lado a lado durante os quase oitocentos anos de dominação muçulmana. O árabe será então um adstrato, nome que se dá a uma língua que permanece falada ao lado de outra, sem se impor ao povo conquistado, mas também sem desaparecer, como ocorreu com o suevo e o visigodo. Esse convívio ficará marcado no léxico das línguas ibéricas, que conservarão muitas palavras de origem árabe, como, no nosso caso, *açafrão, açougue, açude, alambique, alcaide, alcatifa, alcova, alecrim, alface, alfaiate, alfândega, alfazema, alforria, algarismo, algodão, alicate, almofada, almoxarife, alqueire, alvará, armazém, arroz, auge, azar, azeite, azeitona, azulejo, ceroula, chafariz, enxaqueca, fulano, laranja, oxalá, salamaleque, xarope, xaveco, xeque* etc.

Os muçulmanos nunca conseguiram dominar o extremo noroeste da Península Ibérica, onde iam se formar o galego, o leonês e o castelhano, línguas que deixaram documentos escritos. Nas áreas conquistadas pelos mouros, se formaram os chamados dialetos moçárabes ("arabizados"). Era um conjunto de falares românicos utilizados

pelos cristãos sob domínio muçulmano. Esses falares deviam ser bem diferentes das outras línguas românicas que se formariam na parte norte e noroeste da Península, mas infelizmente não dispomos de praticamente nenhum documento confiável sobre como eram esses dialetos. À medida que os exércitos dos reinos cristãos do norte foram expulsando o dominador islâmico, suas línguas também foram descendo na direção sul, passando a ser faladas pelas populações moçárabes das terras reconquistadas. Foi assim que o castelhano se expandiu pela maior parte do que hoje é a Espanha e foi assim também que o galego foi levado cada vez mais para o sul da faixa ocidental da Península, até ser oficializado em 1290 com o nome de português.

As marcantes diferenças que existem entre o português europeu falado no centro-sul do país (de Lisboa para baixo) e o português do norte de Portugal, bem como as diferenças entre o castelhano de Castela e o falado na Andaluzia (centro-sul da Espanha) podem ser resultantes do substrato moçárabe: ao adotarem as línguas dos seus novos senhores cristãos, os falantes de moçárabe teriam transferido para elas hábitos articulatórios típicos de sua língua. O andaluz é tão diferente do castelhano que seus falantes se consideram usuários de uma língua distinta. No entanto, essa hipótese dificilmente será comprovada, já que quase nada conhecemos dos dialetos moçárabes.

Retomando o que vimos até agora, podemos esquematizar a história linguística do galego arcaico da seguinte maneira:



Os mouros serão progressivamente expulsos da Península Ibérica durante o longo período da Reconquista. Os reinos cristãos do norte e do noroeste juntarão forças, receberão ajuda de outros contingentes europeus, até que em 1492, com a queda do reino de

Granada, a Reconquista chegará ao fim. Os reinos de Leão, Castela e Aragão se tornaram um só em 1497, definindo em grandes linhas o território do que hoje é a Espanha.

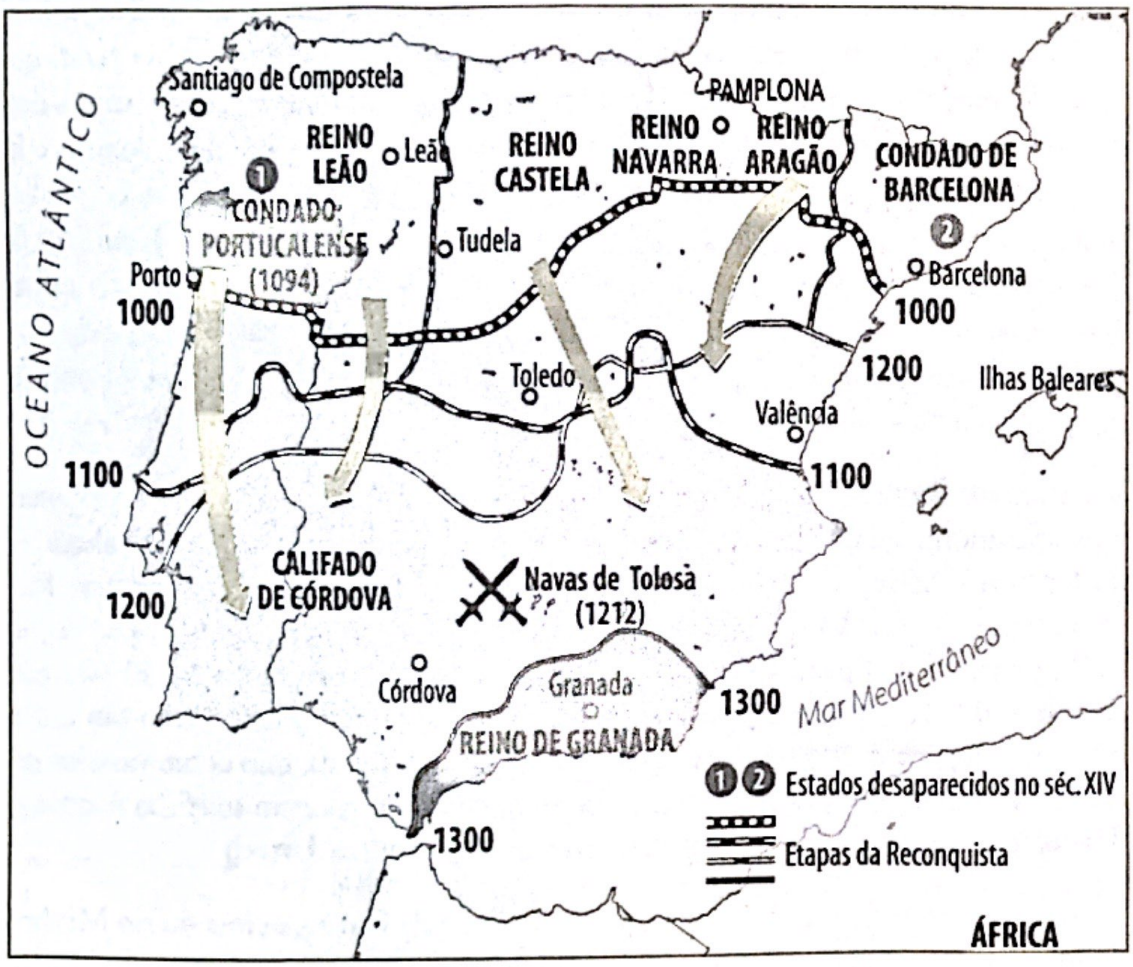
Para incentivar o movimento da Reconquista, os cristãos do noroeste difundiram o mito de que os restos mortais do apóstolo Iago (morto no ano 44) tinham sido descobertos na região, entre 820 e 835, num esquife de pedra supostamente trazido por anjos desde a Palestina logo nas primeiras décadas do cristianismo. A descoberta milagrosa teria ocorrido em Compostela e por sua causa surgiu e se desenvolveu a cidade de Santiago de Compostela (de *Santo Iago, Sant'Iago*: o nome *Tiago*, que se fixou em português, resulta da interpretação equivocada que analisa *Santiago* como formado de *San Tiago*). A cidade se tornou ponto final da peregrinação dos cristãos de toda a Europa que para lá iam venerar as supostas relíquias do santo. E para defender o santuário e o célebre Caminho de Santiago era preciso combater os muçulmanos que dominavam a Península e ameçavam a fé cristã. Com isso, a Reconquista ganha contornos de uma "guerra santa". O mito vai ser ainda reforçado pelas lendas de que o próprio Santiago participava das batalhas ao lado dos cristãos, dando origem à figura de *Santiago Matamoros* ('mata mouros'), muito representada na iconografia religiosa hispânica.

Voltando um pouco nessa história, vamos nos concentrar agora no que aconteceu especificamente na Galiza. Para auxiliar os reinos cristãos a expulsar os muçulmanos do território ibérico, diversos nobres estrangeiros vieram participar das lutas. Em apoio ao rei Afonso VI de Leão e Castela vieram dois nobres franceses: Henrique de Borgonha e seu primo Raimundo de Borgonha. Para recompensar os esforços e as conquistas obtidas pelos primos de Borgonha, o rei concedeu a cada um deles um condado, isto é, um território que pertenceria a cada um, mas como vassalos de Afonso VI. Além disso, promoveu o casamento de Henrique com sua filha ilegítima Teresa (ou Tareja) e o de Raimundo com sua filha legítima Urraca.

O território concedido a Raimundo foi o Condado da Galiza, acima do rio Minho. A Henrique coube, em 1093, o Condado Portucalense, situado entre o rio Minho (que até hoje serve de fronteira entre Portugal e a Galiza) e o rio Mondego, que banha Coimbra. O nome Portucalense deriva da cidade de *Portucale*, a atual cidade do Porto, no norte de Portugal.

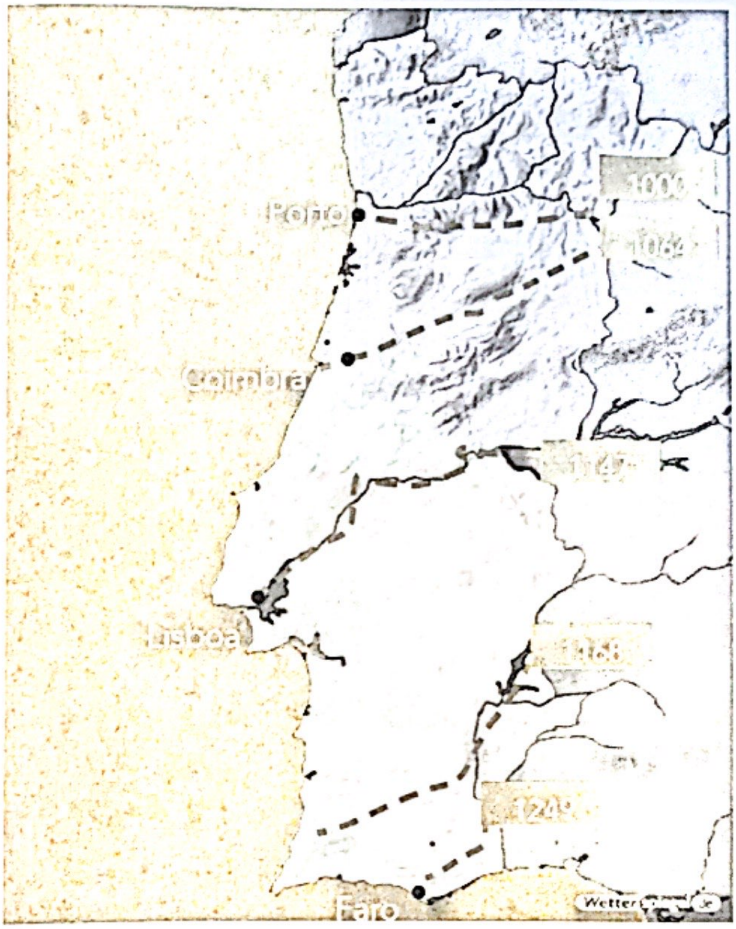
A partir daí começam várias peripécias familiares que marcarão para sempre os destinos separados de galegos e portugueses. Com a morte de seu marido Raimundo e de seu pai Afonso VI, D. Urraca assume o trono como rainha de Leão e Castela, nele incorporando a Galiza. Ela trava guerra contra sua meia-irmã, D. Tareja, que acaba saindo vitoriosa, graças a alianças com nobres poderosos da Galiza. Por sua vez, D. Tareja enfrentará a oposição do próprio filho mais velho, Afonso Henriques, que não quer ver a herança do pai nem fundida com o território galego nem dependente da coroa de Leão e Castela. Na batalha de São Mamede, em 1128, Afonso Henriques derrota as forças da mãe e seus aliados e assume o poder no Condado Portucalense.

Afonso Henriques desejava a autonomia do Condado e faz articulações junto à Santa Sé para que a independência do território seja reconhecida. Enquanto isso, em 1139, as tropas sob seu comando obtêm uma grande vitória sobre os mouros na batalha de Ourique. Segundo a tradição, foi nessa ocasião que os soldados proclamaram Afonso Henriques "rei de Portugal". A independência do novo reino é reconhecida em 1143 pela coroa de Leão e Castela. Em 1179 é a vez do papa Alexandre III aceitar a autonomia do reino de Portugal.



[Mapa 5.7 — Etapas da Reconquista.]

O rei, agora com o título de D. Afonso I, tendo obtido a paz com seus vizinhos cristãos, vai se dedicar a expandir o território de seu reino através de seguidas campanhas militares contra os muçulmanos. Como seus domínios estavam imprensados a norte e a oeste pelos reinos de Leão e Castela, sua única possibilidade de expansão estava no rumo sul. Essas campanhas serão levadas adiante por seus descendentes até que em 1249, com a conquista da cidade de Faro, no extremo sul da faixa ocidental da Península Ibérica, o reino de Portugal adquire o território que desde então permaneceu praticamente intacto. Com isso, Portugal é o único país da Europa ocidental que teve suas fronteiras nacionais definidas bem cedo, ainda na Idade Média.



[Mapa 5.6 – Etapas da Reconquista e formação do reino de Portugal.]

À medida que o reino se expandia para o sul, o centro de gravidade político e administrativo também se deslocava: a capital do novo reino foi estabelecida primeiro em Coimbra e mais tarde, em 1255, já com o país formado, em Lisboa.

O destino da Galiza foi bem diferente. As vicissitudes políticas e familiares levaram o território galego a permanecer integrado à coroa de Leão e Castela. Desde então, a Galiza é uma província da Espanha, nunca se tornou um país independente. Como já vimos, somente a partir do final dos anos 1970 é que o galego se tornará língua oficial da Região Autônoma da Galiza, que, apesar do nome, não é um Estado soberano.

5.6 O PORTUGUÊS NÃO VEM DO LATIM

Nunca é demais lembrar que a questão dos nomes que se dá às línguas escapa da órbita dos especialistas (filólogos, gramáticos, linguistas) e se vincula muito mais a problemáticas de natureza política, cultural, econômica e ideológica.

Quando o conde Afonso Henriques se tornou o primeiro rei de Portugal, a língua que se falava em seu Condado Portucalense com certeza não era diferente da que os habitantes da Galiza falavam. Essa certeza se firma no fato do galego moderno ainda ser muito semelhante ao português europeu e, mais ainda, ao português brasileiro, e

ao fato dos dialetos do norte de Portugal apresentarem muitas semelhanças com os dialetos do sul da Galiza. Outra comprovação é a documentação escrita que sobrevive desde aquelas épocas remotas: produzidos na Galiza ou no território que viria a ser Portugal, esses textos são registrados numa língua que podemos dizer que é uma só.

A esse respeito, vale a pena citar as palavras de Esperança Cardeira, autora portuguesa de um livro sobre a história da sua língua:

> À entrada do ano mil, no Noroeste peninsular, a Galécia Magna, uma região que se estendia da Galiza a Aveiro abarcando, ainda, uma faixa das Astúrias, delimitava já um romance com contornos peculiares. [...] Não é ainda Portugal, não é ainda língua portuguesa. [...] Antes de Portugal, antes do Português, no limiar do século X, já estava constituído um romance [...] (2006: 36-37)

Se "não é ainda Portugal, não é ainda língua portuguesa" e se a própria autora diz que esse romance era falado em toda a Galécia Magna, que outro nome se poderia dar a essa língua que não seja galego? Por que chamar de galego-português uma língua que surgiu "antes de Portugal, antes do Português", como faz, por exemplo, o dicionário *Aurélio*?

> galego-português [...] Língua românica que era falada a N.O. da Península Ibérica ... e cujas fronteiras, a E., eram o leonês e o castelhano e, ao S., abaixo do rio Minho, os dialetos moçárabes que ali se desenvolviam. [Atestada pelo menos desde o séc. VIII, os primeiros documentos nela conhecidos e redigidos por inteiro datam do séc. XI. No séc. XII Portugal, mas não a Galiza, torna-se independente de Leão e se estende para o S., criando-se assim uma fronteira política que, no séc. XIV, já seria também uma fronteira linguística: ao N., o galego, e ao S., o português.]

Esse texto do verbete exibe algumas incoerências, a começar pelo próprio nome dado à língua. Se ela é "atestada pelo menos desde o séc. VIII", quando ainda não existia a entidade política chamada Portugal (e nem mesmo o Contado Portucalense) e se somente no século XIV se estabeleceria uma "fronteira linguística" entre o galego e o português, por que chamar a língua de "galego-português" e não simplesmente de galego, uma vez que a entidade político-geográfica chamada Galécia existia desde a época dos romanos?

É somente no século XIX que os estudiosos portugueses vão criar uma denominação para essa língua única em que foi produzida a rica literatura trovadoresca medieval. Essa denominação será *galego-português*. Esse nome não aparece em nenhum documento antigo. Ao contrário, quando se faz alguma referência à língua da poesia medieval, o nome que aparece é *galego*. O termo *galego-português* foi cunhado como uma espécie de compromisso ideológico entre duas tensões: o reconhecimento de que a língua da poesia medieval era basicamente o galego, e o anseio, de inspiração nacionalista, de incorporar aquela produção literária ao patrimônio cultural do povo português.

De fato, como relata García Turnes (2002: 333), em sua investigação sobre as origens do termo *galego-português* e seu emprego pelos filólogos lusitanos:

> Ao longo de todo o século [XIX] fomos vendo aparecer distintas variantes como *portuguez galliciano, portuguez galleziano, galliziano-português, gallezio-luzitano* ou *gallecio-portuguez*. Não será senão às vésperas do século XX, pela mão de C. Michaëlis, autora que desenvolverá o grosso e o fundamental da sua obra nesta centúria, que encontraremos, alternando com outras formas, o actual *galego-português*.

A pesquisadora argumenta que Carolina Michaëlis será, de fato, a principal responsável pela difusão e, acrescento eu, pela *instituição*, nos estudos históricos, do termo *galego-português*, que se transformará numa *hipóstase* conveniente para designar uma língua que, em tudo, era pura e simplesmente o galego, numa versão literarizada, convencional, cortesã, própria para o uso na poesia lírica e que, com toda probabilidade, nunca foi realmente *falada* como variedade espontânea, materna, de nenhuma comunidade linguística autêntica.

O nome "galego-português" só responde às exigências da filologia lusa do século XIX. Monteagudo (1999: 119-121), depois de apresentar as mais antigas denominações para a língua galega, já a partir do século XIII, afirma justamente que

> a denominação que recebia a língua dos cancioneiros, ao menos nos ambientes eruditos e trovadorescos [...], era a de *galego*, e convida-nos a questionar a moderna denominação, surgida nos ambientes filológicos lusitanos de finais do século XIX, de galego-português.

Por causa de sua localização afastada com relação aos centros políticos, econômicos e culturais do resto da Espanha, a Galiza sempre foi (e até hoje é) a região mais pobre e menos desenvolvida do país. Sua população e sua língua viveram marginalizadas e sempre sofreram forte discriminação por parte dos demais povos que habitam a Espanha (e por parte dos portugueses também: basta ver as acepções que o dicionário *Houaiss* oferece de *galego* como uso característico de Portugal: "indíviduo rude, grosseiro", e de *galegada* como "ato impensado, ignorância, estupidez").

Com isso, o galego ficou durante séculos relegado a uma língua de camponeses, à língua das camadas sociais menos favorecidas da região, sem cultivo literário, enquanto a elite letrada se expressava (e se expressa) em castelhano. Foi necessário esperar o século XIX para se assistir o Rexurdimento (Ressurgimento), um movimento de intelectuais e artistas que desejavam resgatar a língua e a cultura galegas da quase milenar obscuridade em que tinha permanecido. A grande figura de destaque desse movimento foi a poeta Rosalía de Castro (1837-1885), que colocou em versos a fala humilde do povo de sua terra para expressar delicados sentimentos:

Meses do inverno frios
qu'eu amo a todo amar,
meses dos fartos rios
y o doce amor do lar.

Meses das tempestades,
imaxen da delor[1],
que afrixe[2] as mocedades
y as vidas corta en frol[3].

Chegade, e trás d'outono
que as follas fai caer,
nelas deixá que o sono
eu durma do non ser.

E cando o sol fermoso
d'abril torne a sorrir,
que alume o meu reposo,
xa non o meu sofrir.
(Follas Novas, 1880)

1. delor = 'dor'
2. afrixe = 'aflige'
3. frol = 'flor'

A situação marginalizada do galego decerto contribuiu para que os filólogos portugueses do século XIX não sentissem grande estímulo em reconhecer e assumir a evidência de que o português era e é a continuação histórica da língua galega, levada cada vez mais para o sul, à medida que os reis portugueses expandiam seu território. Não seria digno de um povo soberano e conquistador, responsável pelas aventuras marítimas que revelaram o resto do mundo aos europeus, ter como ancestral uma língua de campônios rudes, uma língua sem prestígio. Daí a designação *galego-português*. Por isso não é correto, do ponto de vista histórico-geográfico, afirmar, como fazem todas as obras filológicas e gramáticas históricas, que "o português vem do latim". O português vem do galego — o galego, sim, é que representa a variedade de latim vulgar que se constituiu na *Gallaecia* romana e na Galiza medieval.

No Renascimento, os primeiros gramáticos portugueses trataram de destacar a elegância e a riqueza da língua que então passou a se chamar "português", em contraposição declarada à língua galega. O historiador e gramático Duarte Nunes de Leão, em 1606, escreveu com perspicácia:

> ...as quais ambas [galega e portuguesa] eraõ antigamente quasi hũa mesma, nas palavras & nos diphtongos e na pronunciação que as outras de Hespanha naõ tem. Da qual lingoa Gallega a Portuguesa se aventajou tanto, quanto na copia e na elegancia della vemos. O que se causou por em Portugal haver Reis e corte que he a officina onde os vocabulos se forjão e pulem e onde manão pera os outros homẽs, o que nunqua houve em Galliza...

A presença de reis e de uma Corte é o que permitiu, portanto, à língua portuguesa distinguir-se e separar-se do galego, uma língua que por muitos séculos não será objeto de cultivo literário, relegada aos usos menos nobres, sempre oprimida pelo castelhano centralizador.

Uma voz que se levantou contra essa afirmação da descendência direta do latim para o português, e antes mesmo da criação do nome *galego-português*, foi a do religioso e erudito Martín Sarmiento (1695-1772), nascido na Galiza. Numa de suas muitas obras sobre a língua galega, ele se contrapõe ao gramático e historiador português Duarte Nunes de Leão (1530?-1608) com palavras muito ásperas a esse respeito:

> Duarte Nunes de Leão, português, escreveu um *Compêndio das crônicas portuguesas*, uma *Geografia de Portugal*, outro de *Ortografia portuguesa*, e outro com este título: *Origem da língua portuguesa*. Esse título é semelhante ao que teria um livro disparatado, *Origem da língua andaluza*, no qual não se fizesse menção de que era derivada do idioma castelhano. *Origem da língua portuguesa* sem mencionar que é derivada do dialeto galego é uma fábula, uma quimera, uma nescidade monstruosa, uma fanfarronada e um desvario [...] A linguagem que hoje falam os portugueses e que se acha escrita nos livros até Duarte Nunes nem é língua nem dialeto, mas um subdialeto da língua galega que hoje se fala e antigamente se escrevia. O que devia ter posto Nunes era a origem das vozes mouriscas, das brasileiras, das africanas e das asiáticas que jamais foram galegas, e com as quais se contaminou o primitivo dialeto depois das conquistas dos portugueses rumo ao sul.

De fato, a língua falada no reino de Portugal vai levar 160 anos para receber explicitamente o nome de português. Como em todos os reinos da Europa medieval, o latim literário era a língua oficial da administração e da política, a língua da filosofia e da teologia, mesmo em lugares onde o latim nunca tinha sido falado, em territórios que não tinham feito parte do Império Romano (como a Hungria, por exemplo). O domínio espiritual da Igreja católica romana, que só viria a ser contestado no século XV, com o protestantismo, incentivou esse uso do latim como língua da cultura e da burocracia. Nos países onde o latim vulgar passou a ser falado, as línguas nascidas dele recebiam nomes genéricos como "romance", "romanço", "vulgar", entre outros.

Eram redigidos em latim, portanto, os documentos oficiais, os textos eclesiásticos, as leis etc. No reinado de D. Dinis (iniciado em 1279), de forma pioneira na Europa, a língua falada pelo povo se torna a língua oficial do Estado. Até então era chamada de "língua vulgar". A partir da publicação do decreto de D. Dinis intitulado *Magna Charta Privilegiorum*, a língua passa a se chamar *português* e é adotada oficialmente pelo Estado. D. Dinis, no mesmo documento, funda a Universidade de Coimbra, uma das mais antigas da Europa. Foi provavelmente o primeiro rei de Portugal não analfabeto. Era poeta, tendo escrito diversas cantigas de amor e de amigo. Sua esposa, Isabel de Aragão, célebre por sua piedade e seu amor aos pobres, foi canonizada pela Igreja e se tornou conhecida como Rainha Santa Isabel.

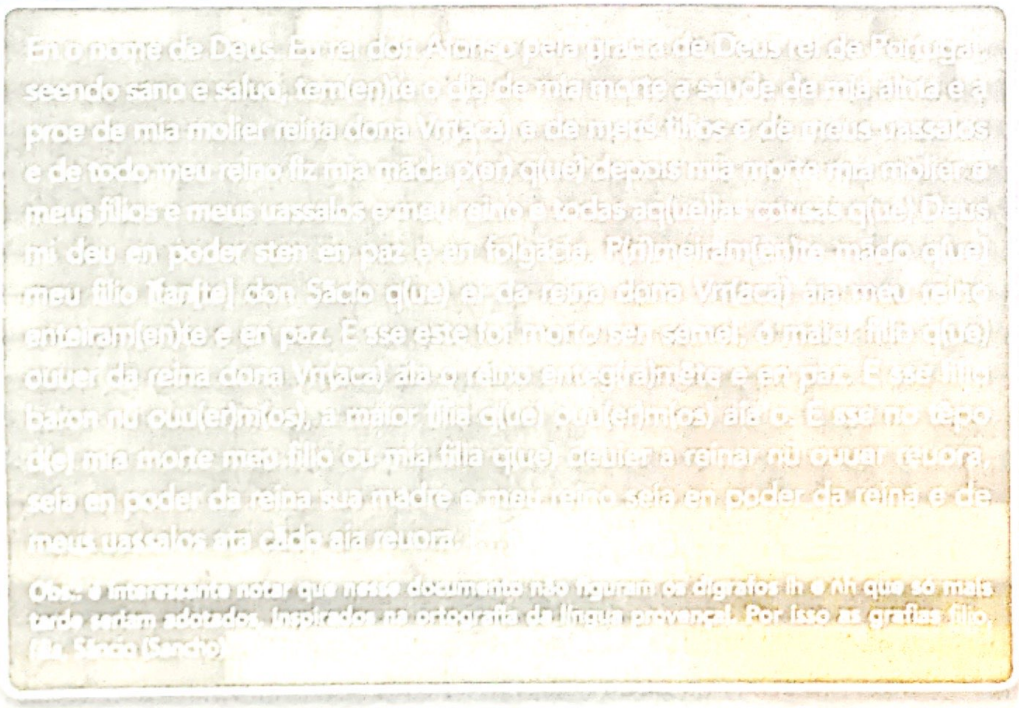
O reinado de D. Dinis ficou marcado por importantes fatos históricos. Foi ele quem mandou plantar os famosos choupais de Coimbra, árvores com que seriam construídas as caravelas que, dois séculos e meio mais tarde, partiriam em busca de novas terras

desconhecidas. Em 1255, a capital do reino é transferida para Lisboa, no centro quase exato do país, de modo a facilitar as comunicações entre o norte e o sul do reino.

A instalação da capital em Lisboa terá profundas consequências sociolinguísticas: a língua, nascida no extremo norte, se difunde até o sul e, nesse processo, vai sofrendo influências e confluências que terminarão por caracterizar os diversos dialetos do português europeu. O português da zona central, onde está Lisboa, aos poucos se diferenciará marcadamente dos dialetos do norte, mais conservadores e mais próximos do galego. Na condição de capital, Lisboa se tornará o centro da norma linguística: a pronúncia, o léxico e a morfossintaxe lisboetas pouco a pouco se tornarão "a língua portuguesa", relegando as demais variedades a usos regionais, sem reconhecimento no sistema educacional e político. Muitas das características linguísticas que os brasileiros atribuem aos portugueses em geral são, de fato, traços próprios do dialeto lisboeta e não ocorrem em outras áreas do país.

Observe-se que D. Dinis é o *sexto* rei de Portugal e que entre a batalha de Ourique (1139), onde Afonso Henriques se proclamou rei, e a publicação da *Magna Charta Privilegiorum* (1290), transcorre mais de século e meio. Por isso, não faz sentido, do ponto de vista cronológico, designar como "português" (mesmo que precedido de "galego", como em *galego-português*) uma língua que já era falada e provavelmente escrita antes mesmo da existência da entidade política que viria a se chamar Portugal.

Um dos textos mais antigos que temos em galego é o testamento do rei D. Afonso II de Portugal, escrito em 1214. Esse rei, ao contrário de seu pai e seu avô, preferiu não entrar em guerra contra os muçulmanos e cuidar mais do fortalecimento interno de seu pequeno reino, que tinha Lisboa como seu limite extremo ao sul. Durante seu reinado, a capital ainda era Coimbra.



En o nome de Deus. Eu rei don Afonso pela gracia de Deus rei de Portugal, seendo sano e salud, tem(en)te o dia de mia morte a saude de mia alma e a proe de mia molier reina dona Vr(aca) e de meus filios e de meus uassalos e de todo meu reino fiz mia mãda p(er) q(ue) depois mia morte mia molier e meus filios e meus uassalos e meu reino e todas aq(ue)las cousas q(ue) Deus mi deu en poder sten en paz e en folgacia. P(r)meiram(en)te mãdo q(ue) meu filio Ilan(te) don Sãcio q(ue) er da reina dona Vr(aca) aia meu reino enteiram(en)te e en paz. E sse este for morto sen semel, o maior filio q(ue) ouuer da reina dona Vr(aca) aia o reino enteiram(en)te e en paz. E sse filio baron nu ouu(er)m(os), a maior filia q(ue) ouu(er)m(os) aia o. E sse no tẽpo d(e) mia morte meu filio ou mia filia q(ue) deuier a reinar nu ouuer reuora, seia en poder da reina sua madre e meu reino seia en poder da reina e de meus uassalos ata cado aia reuora. ...

Obs.: é interessante notar que nesse documento não figuram os dígrafos lh e nh que só mais tarde seriam adotados, inspirados na ortografia da língua provençal. Por isso as grafias filio, (Ilia, Sãcio (Sancho)

[Em o nome de Deus. Eu, rei dom Afonso pela graça de Deus rei de Portugal, sendo são e salvo, temente o dia de minha morte a saúde de minha alma e a prole de minha mulher rainha dona Urraca e de meus filhos e de meus vassalos e de todo meu reino fiz minha manda[1] para que depois de minha morte minha mulher e meus filhos e meus vassalos e meu reino e todas aquelas cousas que Deus me deu em poder estejam em paz e em folgança. Primeiramente mando que meu filho infante dom Sancho que tenho da rainha dona Urraca tenha meu reino inteiramente e em paz. E se este morrer sem semente, o maior filio que tiver da rainha dona Urraca tenha o reino integramente e em paz. E se filho varão não tivermos, a filha mais velha que tivermos tenha-o. E se no tempo de minha morte meu filho ou minha filha que vier a reinar não tiver revora[2], seja em poder da rainha sua mãe e meu reino seja em poder da rainha e de meus vassalos até quando tiver revora.]

(1) manda: testamento, legado.

(2) revora: maioridade.

5.6.1 Características do galego moderno

O galego moderno se diferencia do português europeu e do PB por alguns traços bastante característicos. Por um lado, é uma língua mais conservadora, em que o ditongo escrito *ei* tem sua pronúncia primitiva [eɪ]. O algarismo 2 até hoje se chama *dous* e a palavra *cousa*, que para nós soa antiga, é viva no cotidiano dos galegos. As formas verbais correspondentes ao índice de pessoa *vós* são amplamente usadas e com suas formas mais antigas: *cantades, falades, pedides*.

O galego não desenvolveu a consoante [z], de forma que a palavra *casa* é pronunciada [kaśa] com um [ś] ápico-alveolar semelhante ao do castelhano e ao dos dialetos do norte de Portugal. Do mesmo modo, não desenvolveu a distinção entre [b] e [v], o que também deixa ele mais próximo do castelhano e dos dialetos portugueses do Norte. As terminações *-ane*, *-anu* e *-one* não evoluíram, como em português, para o ditongo [ãʊ̃], o que mais uma vez aproxima o galego do castelhano. Outro desenvolvimento fonético semelhante ao do castelhano foi a consoante interdental [θ], que se escreve *z* ou *c* (diante de *e* e *i*) e totalmente desconhecida dos falantes de português europeu e brasileiro. (É o som que se escreve *th* em inglês, como no verbo *think*, por exemplo.) Em alguns dialetos do galego, no entanto, essa interdental se reduz a [ś]. O galego conserva a pronúncia africada [ʧ], que se escreve *ch*, pronúncia que existiu no português medieval e que permanece no castelhano, representada pelo mesmo dígrafo. Desse modo, *chave* soa para nós, na fala dos galegos, como "tchabe". As vogais postônicas finais *-e* e *-o* não se reduziram no galego a [i] e [u] como no PB, de modo que o que se escreve *ovo* em galego soa para nós como "obo", com dois *o* fechados.

Outro fato interessante é que em galego o feminino de *um* ("um") é *unha*, mas essa grafia não deve nos enganar, porque é o único *nh* que existe na língua (para representar o som do nosso *nh* o galego usa o *ñ*, como o castelhano). O *h* aparece ali para indicar uma velarização, de modo que a palavra se pronuncia ['ũa], em duas sílabas. A ortografia do português, como vimos em [▸127], é responsável pela pronúncia [ũma]: como o masculino se escreve *um* (mas é pronunciado [ũ]), para a formação do feminino se acrescentou um *-a*, embora a pronúncia devesse ser, como em galego, [ũa]. Em textos antigos portugueses aparecem as grafias *ũ* e *ũa*. Ainda mais

interessante é lembrar que em algumas variedades do PB, sobretudo no Nordeste, a pronúncia [ũa], do português medieval, foi conservada.

Por outro lado, o galego seguiu outros desenvolvimentos que representam passos adiante no processo de mudança linguística. Sem dúvida, o mais característico de todos é a transformação da consoante palatal sonora [ʒ] em sua homorgânica surda [ʃ], que na ortografia galega se escreve com *x*. Assim, palavras que em português conservam o [ʒ], como *jantar, jeito, jogo, japonês, gente, geral, geografia, hoje* etc. se escreve em galego *xantar, xeito, xogo, xaponés, xente, xeral, xeografía, hoxe* etc. Mais uma vez podemos lembrar que no Nordeste brasileiro ocorrem as exclamações *Vixe!* (de *Virgem*) e *oxente* (de *ó gente!*).

O galego não conhece a mesóclise: os clíticos usados com verbos no futuro do indicativo e no futuro do pretérito se colocam depois do verbo — observe no pequeno texto abaixo a forma *publicarase* (*publicará* + *se*), que no português clássico se escreveria *publicar-se-á*. A mesóclise, como se sabe, não existe no vernáculo do PB.

Em grandes áreas do território de língua galega, a consoante [g] passou por uma transformação conhecida por *gheada*, em que o [g] se pronuncia como uma aspirada [x], semelhante ao *j* do castelhano em *jamón, Juan, lejos* etc. e que é grafada com o dígrafo *gh*. Assim, *gato* se escreve *ghato* e se pronuncia ['xato], e *galego* se torna *ghalegho* [xa'lexo]. O dígrafo *gh* é usado em holandês para representar o mesmo som, como no nome do pintor Van Gogh.

Abaixo um pequeno texto escrito em galego contemporâneo a respeito do calor que se abateu sobre a Europa no verão de 2010:

> Galicia acaba de pechar un dos meses de agosto máis cálidos e secos das series históricas manexadas por Meteogalicia, axencia de predicción dependente da consellería de Medio Ambiente. Varios valores afástanse nunha primeira análise con carácter provisional — hoxe publicarase o informe mensual definitivo — das cifras medias dos rexistros: os primeiros 30 días do mes (á espera de incluír a xornada de onte) destacan polo escaso número de días de choiva e a reducida cantidade de auga recollida en cada un deles. Así mesmo, as temperaturas máximas e mínimas movéronse este agosto por encima das cifras recollidas para anos anteriores.
> **Galicia Hoxe, 1/9/2010.**

5.7 Brasil: línguas indígenas

Em 1498, com Vasco da Gama, é a vez dos portugueses seguirem o exemplo dos romanos e saírem pelo mundo invadindo novas terras, conquistando, escravizando e dizimando povos e impondo sua língua e sua religião. Esse empreendimento colonial e imperial vai se desenrolar primeiro na Índia e na África e, em 1500, na América do Sul, com a invasão e ocupação portuguesa das terras que mais tarde formariam o Brasil.

Não se sabe dizer ao certo quantas eram as nações indígenas que habitavam o território brasileiro no início do século XVI, nem qual era a população total. Também não é possível determinar com exatidão o número de línguas diferentes empregadas por esses povos. Os números variam entre 1 e 6 milhões de habitantes e 300 línguas. Hoje, os descendentes desses povos somam menos de 300 mil, falando cerca de 180 línguas, muitas delas empregadas por um número extremamente reduzido de pessoas, o que sinaliza sua provável extinção futura.

A chegada no Brasil de padres jesuítas com a missão de cristianizar os índios deu início a um interessante fenômeno linguístico que foi o surgimento das chamadas línguas gerais. Essas línguas gerais são, de fato, línguas do grupo tupi que os jesuítas empregaram para se comunicar com os indígenas, transmitir a eles as crenças europeias, convencê-los dos valores supostamente superiores da civilização cristã. É precisamente de um jesuíta, José de Anchieta, a primeira gramática escrita de uma língua indígena: *Artes de gramática da língua mais usada na costa do Brasil* (1595), uma descrição da língua tupi. Anchieta, como se sabe, também escreveu peças de teatro e poemas nessa língua.

Duas línguas gerais se desenvolveram assim: a língua geral amazônica (ou nheengatu, "língua boa, bonita"), empregada na porção norte da colônia, e a língua geral paulista, empregada inicialmente em São Paulo e posteriormente difundida para outras regiões através da ação dos bandeirantes, que penetravam nos sertões interiores para capturar índios e escravizá-los. Essas línguas eram empregadas pelos padres, pelos índios, por muitos dos portugueses que viviam no Brasil e por seus descendentes mestiços, uma vez que, na falta de mulheres brancas na colônia, os casamentos de brancos com índias eram uma prática muito comum. Com isso, por um longo período inicial da nossa história colonial, o português foi língua minoritária, uma vez que a maioria da população era composta de indígenas e mestiços, que aprendiam a língua da mãe. É muito conhecida e citada a seguinte passagem de uma carta do padre Antônio Vieira, jesuíta, que viveu boa parte de sua vida no Brasil, no século XVII:

> É certo que as famílias dos Portugueses e Índios de S. Paulo estão tão ligadas hoje umas com as outras, que as mulheres e os filhos se criam mestiça e domesticamente, e a língua que nas ditas famílias se fala é a dos Índios, e a portuguesa a vão os meninos aprender à escola. (Carta do Pe. Antônio Vieira, 1694)

Poucos anos depois da carta de Vieira, o governador de São Paulo, Artur de Sá e Meneses, escrevia em 1698 ao rei de Portugal solicitando que os padres enviados para aquela região fossem conhecedores da língua geral porque, em suas palavras,

> [...] a maior parte daquela gente se não explica em outro idioma, e principalmente o sexo feminino e todos os servos, e desta falta se experimenta irreparável perda, como hoje se vê em São Paulo com o novo Vigário que veio provido naquela Igreja, o qual há mister quem o interprete.

O religioso Félix de Azara, exercendo suas funções no Paraguai, observava, num texto escrito cem anos depois de Vieira e Meneses, que ali também predominava a língua guarani em detrimento do espanhol e comparava a situação com a de São Paulo (província que naquela época abarcava Minas Gerais, Mato Grosso e Goiás):

> Lo mismo ha sucedido exactamente en la imensa provincia de San Pablo, donde los portugueses, habiendo olvidado su idioma, no hablan sino el guaraní.

De fato, o mesmo uso sistemático de uma língua geral para catequizar os indígenas foi feito pelos jesuítas no Paraguai com o *guarani*. Até hoje, a população paraguaia é majoritariamente bilíngue, empregando o espanhol paraguaio e o guarani (chamado *abanheenga*, "língua de homem, língua de gente"), sendo que a maioria dos falantes monolíngues só conhecem o guarani. O mesmo, talvez, poderia ter ocorrido no Brasil se, entre outras coisas, a Coroa portuguesa não tivesse proibido explicitamente o emprego de qualquer outra língua no ensino que não o português. Essa proibição fazia parte de uma política maior, que incluiu também a expulsão dos jesuítas das terras brasileiras. Isso ocorreu em 1757, durante o governo do Marquês de Pombal como primeiro-ministro de Portugal.

É claro que uma língua não deixa de ser falada por decreto, por mais que isso tenha algum impacto real sobre a vida dos falantes. Os principais fatores históricos que determinaram o desaparecimento da língua geral paulista são de ordem econômica. Com o descobrimento do ouro nas Minas Gerais, no final do século XVII, a captura de índios para a escravização declinou em favor do envio de grandes contingentes de escravos negros para as minas. Ocorreu também uma grande migração de portugueses (cerca de 300 mil) para o Brasil, desejosos de enriquecer com a exploração do ouro. Isso decerto favoreceu a expansão do uso do português, tanto em sua versão europeia quanto em sua versão brasileira, que já ia se constituindo graças sobretudo aos escravos. O deslocamento do principal polo da economia colonial também provocou o deslocamento da capital da colônia de Salvador para o Rio de Janeiro em 1765, porto que se firmou como o principal escoadouro do ouro do interior.

Assim, a língua geral paulista foi aos poucos deixando de ser falada, até extinguir-se por completo. Ao que tudo indica, já no final do século XVIII, o português predominava na parte meridional da colônia. Por seu turno, a língua geral amazônica, apesar de proibida, continuou a ser falada, embora cada vez por menos gente. Hoje em dia, com o nome de *nheengatu*, ela é empregada pela população que vive às margens do rio Negro, no estado do Amazonas. Na cidade de São Gabriel da Cachoeira, banhada pelo Negro, o nheengatu se tornou língua cooficial do município, em 2002, ao lado do português, do baníua e do tucano.

As populações indígenas brasileiras foram sistematicamente massacradas ao longo de todo o período colonial e também depois da Independência. Até hoje, os índios brasileiros são alvo de perseguição e assassinato por parte de latifundiários e de outros

interessados em explorar as terras indígenas. Os índios que foram viver no meio urbano compõem o segmento mais miserável da população, vivendo em condições precárias e de grande penúria. O fenômeno do suicídio de índios é registrado em diversas regiões brasileiras: em 2009, por exemplo, ocorreu um suicídio a cada dez dias no Mato Grosso do Sul. A taxa de mortalidade de bebês indígenas é o dobro da média nacional.

Apesar do grande número de línguas indígenas, o tupi foi a que mais contribuições deu ao PB, sobretudo no léxico, por causa principalmente da ação dos bandeirantes paulistas, que denominavam com palavras tupis os locais e acidentes geográficos que encontravam em seu caminho sertão adentro. Por essa razão, é comum encontrar toponímia de origem tupi mesmo onde nunca houve uma população tupi original. Se calcula em torno de dez mil os vocábulos de origem tupi empregados na nossa língua. A maioria deles se refere à flora e à fauna e a características ecológicas, geológicas e hidrológicas do Brasil. Existem também outras palavras de uso muito frequente no PB que provêm do tupi: *arapuca, arataca, biboca, caatinga, caiçara, caipira, capão, capim, capoeira, catapora, coroca, cuia, cuíca, guri, jururu, mingau, muquirana, mutirão, nhenhenhém, paçoca, peba, pereba, peteca, pindaíba, pipoca, pito, pixaim, sapecar, sarará, sururu, tapera, taquara, taturana, tipoia, tiririca, tocaia* etc.

TERMO TUPI	SIGNIFICADO	EXEMPLOS
açu, uaçu, guaçu	grande	Iguaçu (PR): "água grande"; Açu (RN); Paraguaçu (BA): "rio grande"; Igaraçu (PE): "canoa grande"
aíba	ruim, mau, imprestável	Paraíba: "rio imprestável [para a navegação]"; Parnaíba (PI): (idem)
anhã, anhanga	gênio, espírito, demônio	Anhangabaú (SP): "rio do malefício"
beraba	brilhante	Uberaba (MG): "rio brilhante"; Itaberaba (BA): "pedra brilhante"
caá, ca	mato, planta, folha	Cajuru (SP): "boca do mato", Caatiba (BA): "matagal"; Caapiranga (AM): "mato vermelho"
catu	bom, bonito	Catu (BA); Catuípe (RS): "na água boa, no rio bom"
coaraci, guaraci	sol	Coaraci (BA); Guaraci (PR); Guaraciaba (MG): "cabelo do sol", isto é, "louro"
curi	araucária	Curitiba: "pinheiral"
guá, goá	vale, baía, golfo	Paranaguá (PR): "vale do rio", isto é, "baía"; Guanabara (RJ): (idem); Jaraguá (SP): "senhor do vale"
guará	garça	Guaratinguetá (SP): "abundância de garças brancas"; Guarapiranga (SP): "garça vermelha"; Guarabira (PB): "garça empinada"

i, ig, u	água, rio	Tatuí (SP): "rio do tatu"; Jacareí (SP): "rio do jacaré"; Jaguari (RS): "rio da onça"; Jundiaí (SP): "rio do jundiá [peixe]"
ibi	terra, solo	Ibiúna (SP): "terra preta"; Ibitinga (SP): "terra branca"; Ibipitanga (BA): "terra vermelha";
ibitu, botu, votu	vento, ar, nuvem	Botucatu (SP): "bons ares"; Votuporanga (SP): "bons ares, ventos suaves"; Botuporã (BA): "idem"
igara	canoa	Igaraçu (PE): "canoa grande"; Igarapava (SP): "porto das canoas"; Igaratinga (MG): "canoa branca"
itá	pedra	Itacoatiara (AM): "pedra pintada"; Itabuna (BA): "lugar da pedra preta"; Itaú
itu	salto, cachoeira	Itu (SP); Ituverava (SP): "cachoeira brilhante"; Ituberá (BA): (idem); Itumbiara (GO): "caminho da cachoeira"
juca	podre, estagnado	Tijuca (RJ): "água podre", isto é, "charco"; Ipojuca (PE): "pântano"
maracá	chocalho	Itamaracá (PE): "pedra que chocalha"
mirim	pequeno	Moji-Mirim (SP): "pequeno rio que serpenteia"
panema	ruim, imprestável, podre	Ipanema (RJ): "água imprestável"; Upanema (RN): (idem); Paranapanema (SP): "rio imprestável", isto é, "sem peixes"
pará, paraná	rio, mar	Pará; Paraná; Paraíba; Ji-Paraná (RO): "rio-machado"
peba, peva	chato, liso, plano	Itapeva (SP): "pedra chata"
pindá	anzol	Pindamonhangaba (SP): "lugar de fazer anzol"; Pindaí (BA): "rio do anzol"
pirá	peixe	Piracicaba (SP): "onde o peixe para"; Pirapora (MG): "salto do peixe"; Piraçununga (SP): "barulho do peixe"; Piraí (RJ): "rio do peixe"
piranga	vermelho	Ipiranga (SP/PR): "rio vermelho"; Itapiranga (AM/SC): "pedra vermelha"; Caapiranga (AM): "mato vermelho"
poranga, porã	bonito	Poranga (CE); Itaporanga (PB): "pedra bonita"; Ituporanga (SC): "cachoeira bonita"
soroca	fenda, rasgão	Sorocaba (SP): "terra rasgada"
taba	aldeia	Tabatinga (AM): "aldeia branca"; Tabapuã (SP): "aldeia reunida"; Tabaporã (MT): "aldeia bonita"

taguá, tauá	barro, argila	Taguatinga (DF): "barro branco"; Taguaí (SP): "rio do barro"
taquá, taquara	taquara	Taquara (RS); Taquari (RS): "rio das taquaras"; Taquaratinga (SP): "taquara branca"
tiba, tuba, tuva, nduva, ndiva, tiua, tiva	abundância (sufixo coletivo)	Ubatuba (SP): "muitas canoas"; Itatiba (SP): "muitas pedras"; Caatiba (BA): "matagal"; Taquarituba (SP): "taquaral";
tinga, tininga	branco (ou, por extensão, 'seco')	Piratininga (SP): "peixe branco", isto é, "seco"; Itatinga (SP): "pedra branca"; Itinga (MA): "rio branco"; Upatininga (PE): "lagoa seca"
una, ú	preto	Una (BA); Itaúna (MG): "pedra preta"; Unaí (MG): "rio preto"; Jaguariúna (SP): "rio preto da onça"

[Alguns termos de origem tupi frequentes na toponímia brasileira.]

Os pesquisadores ainda não conseguiram chegar a um consenso a respeito de alguma influência das línguas indígenas sobre o PB que não seja a contribuição ao léxico. Por exemplo, há quem atribua as pronúncias das consoantes [ʃ] e [ʒ] como [ʧ] e [ʤ] no chamado dialeto cuiabano como uma característica herdada dos índios bororos que habitavam a região. Esses índios falavam uma língua do grupo macro-jê, diferente do tupi. Como se sabe, palavras como *chuva* e *caixa* soam, nesse dialeto, como ['ʧuva] e ['kaʧa], assim como *gente* e *laranja* soam como ['ʤẽte] e [la'rãʤa].

Em seu livro clássico, *O dialeto caipira* (1920), Amadeu Amaral informa que também no interior de São Paulo, naquela época, existiam as pronúncias [ʧ] e [ʤ].

Outro aspecto fonético que algumas pessoas propõem atribuir a um substrato tupi é o chamado "R caipira" (retroflexo), tão característico de uma ampla área do centro-sul do Brasil. No alfabeto fonético seu símbolo é [ɹ]. A justificativa para tal atribuição é o fato dessa área corresponder em grandes linhas ao território onde era falada ou para onde foi levada a língua geral paulista, que era tupi. Além disso, o [ɹ] também ocorre no espanhol do Paraguai, país onde a maioria da população emprega o guarani, língua muito próxima do tupi. No espanhol paraguaio, contudo, esse [ɹ] soa mais brando do que o do dialeto caipira. Amadeu Amaral, no livro já citado, ao tratar do [ɹ] diz: "É, muito provavelmente, o mesmo *r* brando dos autóctones", isto é, dos índios.

No tupi e no guarani o padrão silábico é CVCV, consoante-vogal-consoante-vogal, de modo que não existem encontros consonantais (a palavra *cruz* se transformou em *curuçá* na fala tupi), o que também é característico de outras línguas, como o japonês (onde *Brasil* se diz *Burajiru*). Essas línguas também não têm sílaba travada por consoante (como em *carta, pasto, delta*). É provável que, tendo de pronunciar palavras do português com travamento silábico com [r], os falantes da língua geral articulassem esse [r] com uma retroflexão que fazia a consoante se transformar praticamente numa semivogal, tornando a palavra mais próxima de seu padrão silábico: porta ['pɔrta] (CVCCV) → ['pɔɹta] (CVVCV).

Amadeu Amaral também explica que na fala caipira o grupo *gu* + vogal tende a ser pronunciado como se não houvesse o g: "A explosiva gutural *gh* tem uma tonalidade especial, sobretudo antes dos semiditongos cuja prepositiva é *u*, caso em que frequentemente se vocaliza: *áu-ua* = água, *léu-ua* = légua". Em tupi, segundo alguns pesquisadores, não se pode dizer que existisse a consoante [g]: o que ocorria era uma forte velarização da semivogal [ʊ] que, aos ouvidos dos portugueses, soava como o *gu* de *água*. Isso explica por que, nos topônimos brasileiros, é comum ocorrer o elemento *-guaçu*, do tupi *wasu*, 'grande'.

Fenômeno semelhante ocorreu quando as línguas românicas adotaram palavras germânicas com a semivogal [ʊ] também muito velarizada: *wardon* > *guardar*; *werra* > *guerra*; *winda* > *guindar*; *Wilhelm* > *Guilherme* etc. Para ouvidos brasileiros, a palavra *huevo* ('ovo') em muitas variedades do espanhol americano soa como *güebo*.

Segundo Amadeu Amaral, o dialeto caipira não apresentava distinção entre [b] e [v], havendo muito intercâmbio de uma consoante pela outra. Ora, o tupi apresentava uma única consoante [β], diferente tanto do [b] quanto do [v] do português. Isso explica por que, na toponímia brasileira, ocorrem hesitações como *Itu-verava* e *U-beraba*, nomes formados com o elemento *βeraβa*, "brilhante": "cachoeira brilhante" (Ituverava) e "água brilhante" (Uberaba).

[Mapa 5.9 - Área de maior ocorrência do "R caipira".]

5.8 Brasil: as línguas africanas

Os portugueses, além do pioneirismo nas navegações que revelaram aos europeus o resto do planeta, também foram responsáveis pela reintrodução na economia moderna de um sistema de produção que tinha desaparecido na Europa cerca de mil anos antes: a escravidão. Já em 1552, 10% da população de Lisboa era composta de escravos africanos.

O sequestro e a coisificação de seres humanos provenientes do continente africano representam uma abominação só comparável, talvez, ao massacre dos armênios perpetrado pelo governo turco entre 1915 e 1916 (mais de um milhão de mortos), ao extermínio planejado dos judeus e outras etnias pelo regime nazista na Alemanha entre 1939 e 1945 (mais de seis milhões de vítimas, ¾ da população judaica da Europa), ao extermínio metódico da classe média camponesa (os *kulaks*) sob o regime de Stalin na União Soviética (mais de cinco milhões de mortos) ou ao massacre da população cambojana pelo regime de Pol Pot a partir de 1975 (cerca de 1,7 milhão de mortos).

Esses eventos, porém, por mais que suscitem nosso horror, tiveram uma duração pontual ou limitada a alguns poucos anos. A escravização dos negros africanos, entretanto, se prolongou por quatrocentos anos e atingiu, segundo os estudiosos, dez milhões de seres humanos sequestrados de sua terra natal, sem contar seus descendentes nascidos no cativeiro. Desses africanos transportados para o continente americano, 40% foram trazidos para o Brasil, que é assim o país que mais recebeu escravos negros na história. Foi também o último país do mundo a abolir a escravidão negra, em 1888, de modo que em mais de 75% dos pouco mais de quinhentos anos de nossa história oficial predominou no país a prática escravagista. E nem é preciso lembrar que, mesmo hoje em dia, a situação dos não-brancos brasileiros é degradante, devido ao impregnado racismo que caracteriza nossas relações sociais: 70% dos analfabetos brasileiros são negros e mestiços, e 65% da população carcerária do país é composta de não-brancos. Embora representem metade da população brasileira, somente 14% deles têm nível superior, e a presença de não-brancos no Congresso Nacional é de apenas 13,1% no ano de 2010.

Na Paraíba, morrem 1.083% mais pretos. Em Alagoas, 974%. E na Bahia dos blocos de carnaval, 439,8%
[*Carta Capital*, 9 mar 2011, p. 25.]

A escravidão não se extinguiu completamente em nosso país. Entre 1995 e 2005, foram libertados cerca de 18.000 trabalhadores que viviam em regime de escravidão em grandes propriedades rurais brasileiras. É cabível imaginar que outros tantos ainda vivam em condições semelhantes. Na região Norte, é muito comum que pais pobres entreguem suas filhas para serem "criadas" por famílias de classe média e alta nas cidades. Essas meninas se tornam praticamente escravas, tendo de exercer todo o trabalho doméstico, sem salário, sem direitos civis, sem educação formal, frequentemente submetidas a maus-tratos e a abusos sexuais por parte dos patrões e de seus filhos[1]. O tratamento dispensado em geral às empregadas domésticas — a começar pelo tamanho exíguo das "dependências de empregada" nos imóveis de hoje —, muitas das quais não têm direitos trabalhistas e cumprem jornadas de trabalho de doze horas diárias em seis dias da semana (para não mencionar as que "dormem no emprego") é o prosseguimento, nos dias de hoje, da divisão entre casa-grande e senzala do período escravagista. Em 2009, por exemplo, no Distrito Federal (unidade da federação com a maior renda *per capita* do Brasil), 81% das empregadas domésticas eram negras; dessas, 50,8% eram analfabetas ou não tinham concluído o ensino fundamental, contra 9,7% das não-negras[2].

Para o estudo da formação do PB, o conhecimento da história da escravidão é fundamental e incontornável. O elemento africano sem dúvida é responsável por muitas das características gramaticais específicas do PB, características que tornam o PB diferente não só do português europeu como também das demais línguas da família românica. Infelizmente, porém, o profundo racismo entranhado na pequena parcela dominante de nossa sociedade tem levado ao descaso, ao menosprezo ou ao total obscurecimento do profundo impacto que os africanos e seus descendentes tiveram sobre o PB. Essa invisibilidade do negro em nossa história linguística e sociolinguística só muito recentemente começou a ser denunciada e combatida. É certo que na década de 1950 o filólogo Serafim da Silva Neto apontava para a importância dos africanos na constituição do PB. No entanto, as escolas de pensamento linguístico mais recentes, centradas em estudar a língua "de dentro para dentro", sem consideração dos fatores políticos, culturais, econômicos, sociais etc. que orientam e desorientam a vida de qualquer comunidade humana, se esforçam por explicar as características próprias do PB pelo recurso à tese da "deriva secular", isto é, analisando as transformações ocorridas em nossa língua como meras "evoluções" que já estavam "embutidas" ou "em germe" no sistema da língua e que, em solo brasileiro, simplesmente desabrocharam. É certo

---

1 Um dos casos mais chocantes registrados ocorreu em 2005, em Belém. Naquele ano, Marielma de Jesus Sampaio, de 11 anos, foi torturada, violentada, espancada e morta pelos patrões Roberta Sandreli Rolim e Ronivaldo Furtado. Antes de ser assassinada, a menina foi humilhada, torturada e mantida em cárcere privado durante quatro meses.

2 www.mte.gov.br/observatorio/PED_DF_Mulher_2010.pdf (acesso em 1/9/2010).

que existem tendências universais de mudança decorrentes de fatores cognitivos, mas também é certo que não se pode desconsiderar o impacto profundo dos africanos em todos os aspectos da vida brasileira e prioritariamente na nossa língua,. acelerando as tendências à mudança.

Ora, uma das forças que impulsionam a mudança das línguas, como sabemos, é o contato linguístico. A adoção de uma língua estrangeira por toda uma população conquistada ou escravizada implica, inevitavelmente, na alteração e transformação da gramática dessa língua imposta. Se as línguas românicas são tão parecidas umas com as outras mas, ao mesmo tempo, são tão diferentes que seus falantes não se compreendem mutuamente, é porque, nos diversos territórios conquistados pelos romanos, o latim vulgar foi imposto a falantes de línguas diferentes. Foi o *contato* do latim com as línguas célticas, por exemplo, faladas no noroeste da Península Ibérica que fez nascer o galego. Na mesma Península e bem ao lado surgiu também o castelhano, que guarda muitos aspectos semelhantes ao galego (o abrandamento das consoantes intervocálicas, conforme se examinou em [▸137]), mas que também apresenta diferenças, por exemplo, no que ocorreu com determinadas vogais, que em castelhano se ditongaram (*fuego, piedra, rueda, cierto*) mas não em galego (*fogo, pedra, roda, certo*). Se a Itália surpreende as pessoas pela grande quantidade de línguas diferentes faladas em seu território (e que são designadas, um tanto pejorativamente, como *dialetos*) é porque o latim, que era a língua de uma pequena região central da Península, foi imposto a povos que falavam outras línguas (etrusco, osco, úmbrio, sabélio, grego etc.) [▸202]. Do contato dos falantes dessas outras línguas com o latim é que surgiram os diferentes "dialetos" italianos, muitos deles tão diferentes entre si quanto o PB do italiano.

É impossível desconsiderar o impacto dos africanos sobre o PB quando as fontes históricas nos informam que durante a maior parte da história colonial do Brasil a população branca era extremamente minoritária, enquanto a população negra e mestiça — sobretudo depois do quase extermínio das nações indígenas — representava a ampla maioria. Na época da Independência (1822), mais de ⅔ da população brasileira era composta de negros e mestiços. Foi necessário um plano governamental de explícito "branqueamento da população", com estímulo à imigração europeia (alemães e italianos sobretudo), para que a balança racial da população começasse a se equilibrar. Mesmo assim, o Brasil ainda é a segunda maior nação negra do mundo, superado apenas pela Nigéria.

As condições de vida e de trabalho dos escravos, inicialmente nas lavouras de cana de açúcar do Nordeste, mais tarde, no século XVIII, nas extrações de ouro de Minas Gerais e, no século XIX, nas fazendas de café do Vale do Paraíba, impediam que eles tivessem um contato mais intenso com a língua dos seus senhores. Por isso, através de um processo chamado de transmissão linguística irregular (termo proposto pelo linguista brasileiro Dante Lucchesi), eles aprendiam o português de forma fragmen-

tada e assistemática. Esse português, então, era reestruturado, regramaticalizado pelas gerações seguintes para que pudesse atender a todas as necessidades de interação verbal de seus falantes. Com isso, foi se formando ao longo do tempo o vernáculo brasileiro [▶102], isto é, a língua da imensa maioria da nossa população, com uma gramática muito diferente da gramática das variedades urbanas de prestígio. Antes de discutirmos mais a fundo as consequências do surgimento desse vernáculo, vamos tratar das línguas africanas trazidas para o Brasil pelos escravos.

A maioria dos escravos vindos para o Brasil falavam línguas da família banta, principalmente o quimbundo, o umbundo e o quicongo, muito aparentadas entre si e que até hoje são amplamente utilizadas em Angola e em outros países vizinhos. A contribuição lexical dessas línguas ao PB é notável: utilizamos diariamente uma grande quantidade de palavras de origem banta sem nos darmos conta disso.

Algumas palavras de origem banta do léxico do PB:

agogô	babá	babaca	babaquara	babau	bafafá	bagunça
balangandã	bamba	banguê	banguela	banzé	banzo	baticum
batuque	beleléu	berimbau	biboca	bidu	bimba	binga
boboca	bololô	bozó	brucutu	bumba	bunda	cabaça
cabaço	caçamba	cachaça	cachimbo	cacimba	caçuá	caçula
cafofo	cafuçu	cafundó	cafuné	cafungar	cafuringa	cafuzo
calango	calombo	calundu	calunga	cambada	cambembe	camumbembe
camundongo	candango	candomblé	candonga	canga	cangaço	cangalha
canjica	capanga	capenga	carimbo	caruru	catimba	catina
catota	caxambu	caxinguelê	caxixi	caxumba	cochilar	conga
coroca	cotó	cuca	cuíca	curiango	curinga	cutuba
dendê	dengo	dunga	encabular	encafifar	fiofó	fuá
fubá	fungar	fuxico	fuzuê	galalau	ganzá	ginga
iaiá	imbondo	inhaca	jagunço	jiló	lelé	lengalenga
lundu	maconha	maculelê	macumba	mafuá	malungo	mambembe
mandinga	maracutaia	marimbondo	maromba	maxixe	meganha	miçanga
milonga	minhoca	mocambo	mocó	mocotó	molambo	moleque
mondrongo	monjolo	moqueca	moringa	muamba	mucama	munganga
mungunzá	murundu	muvuca	muxiba	muxoxo	olelê, olalá	pitoco
pongar	quenga	quiabo	quibebe	quilombo	quindim	quitanda
quitute	quixó	quizila	quizumba	sacana	samba	senzala
songamonga	sunga	tabaca	tanga	tipoia	titica	tribufu
tunga	tutu	umbanda	vatapá	xibiu	xibungo	ximbica
xingar	xinxim	xoxota	zabumba	zangar	zanzar	ziquizira

[Fonte: Yeda Pessoa de Castro, *Falares africanos na Bahia: um vocabulário afro-brasileiro.* Rio de Janeiro: ABL/Topbooks, 2ª ed., 2006.]

Além das línguas da família banta, os escravos também trouxeram para o Brasil línguas do grupo oeste-africano, entre as quais o iorubá. Por causa do prestígio adquirido pelos cultos religiosos do candomblé, trazidos para o Brasil, e mais especificamente para a Bahia, muita gente acredita, erroneamente, que é o iorubá a língua que melhor representa a africanidade do nosso povo. Isso leva muitas pessoas a querer estudar essa língua, na crença de que assim se aproximariam mais das nossas raízes africanas. No entanto, a contribuição do iorubá ao PB é bastante restrita, quase toda pertencente ao domínio da religião. De fato, 71% das contribuições lexicais africanas ao PB são de origem banta. Os escravos bantos foram trazidos para o Brasil muito antes dos iorubás e foram dispersados por quase todas as regiões do país. Os iorubás chegaram mais tarde e foram concentrados principalmente no Recôncavo Baiano e em outros pontos (ver mapa abaixo). A simples observação da tabela acima comprova a importância considerável dos bantos para a nossa cultura, nossas relações afetivas, nossos modos de agir e ver o mundo. E sem dúvida é de maioria banta a porcentagem do nosso DNA de procedência africana, conforme demonstraram os pesquisadores em genética. É raro encontrar uma brasileira ou um brasileiro, mesmo de cabelos louros e olhos azuis, que não tenha em seu código genético um componente de ascendência africana.

[Mapa 5.10 – Distribuição dos escravos africanos nas regiões brasileiras.]

A pesquisadora Yeda Pessoa de Castro considera que, além do léxico, as línguas africanas também contribuíram para formar o que ela chama de matriz africana do

português brasileiro. Segundo a autora, é possível identificar no PB, em suas variedades "populares", outras influências dos idiomas africanos:

- a tendência do falante brasileiro em omitir as consoantes finais de palavras ou transformá-las em vogais (*fazê, dizê, Brasiu, sinhô, mulhé*), que coincide com a estrutura silábica das palavras em banto e iorubá, que nunca terminam em consoante;
- nessas línguas não ocorrem encontros consonantais, o que explicaria a ocorrência de formas como *saravá* ("salvar"), *fulô* ("flor"), *terém* ("trem") etc.;
- a concordância de número presente apenas no artigo ou outro determinante (*as casa, esses menino, uns livro, dois carro*), tal como em banto, onde o plural dos nomes é marcado por meio de prefixos.

Tem autores que postulam a hipótese de que as pronúncias palatalizadas de [d] e [t] diante de [i], amplamente difundida no território brasileiro — ['ʤia] (*dia*), ['ʧia] (*tia*) — possam ser resultantes de influência do substrato africano. De fato, no português angolano e santomense essas pronúncias também ocorrem.

Um testemunho da antiguidade dos principais traços do vernáculo brasileiro, resultantes do contato linguístico dos escravos africanos com o português, são as recomendações que o bispo Azeredo Coutinho fazia em 1798 às professoras de um internato que recebia meninas vindas do interior da província de Pernambuco. Em sua carta, o bispo aponta alguns "vícios" que era necessário corrigir:

- inversão de "letras": *breço* em vez de *berço*;
- supressão de "letras": *teado* em vez de *telhado* (mais provavelmente deveria ser *teiado*);
- supressão da marca de plural: *muitas flore* em vez de *muitas flores* (o bispo não observava que havia marca de plural, porém só no determinante).

Convém notar que as meninas eram filhas de senhores brancos, proprietários rurais. Sem dúvida, a convivência com os escravos domésticos, e principalmente com as amas, influenciava os modos de falar das crianças da casa-grande.

As características apontadas por Yeda P. Castro podem ser facilmente verificadas numa rápida descrição do quimbundo, uma das línguas do grupo banto falada por um grande contingente de escravos trazidos para o Brasil. Em 1697 foi publicada em Lisboa a *Arte da língua de Angola*, escrita pelo padre jesuíta Pedro Dias. O que o autor chamava de "língua de Angola" é, na verdade, o quimbundo. Vamos ver algumas de suas observações:

> 1. "Todos os nomes, & verbos acabaõ em as vogaes, a, e, i, o, u [...]" (Dias, 1697: 2)
>
> O quimbundo tem uma estrutura silábica do tipo CV.CV (consoante-vogal-consoante-vogal), em que algumas consoantes são precedidas de uma nasalidade (*nvula*, 'chuva'). Isso explicaria a tendência do PB, sobretudo das variedades rurais e rurbanas, a eliminar as consoantes em final de palavra (*fazê, cantá, amô, sinhô*) e a romper os encontros consonantais pela inserção de uma vogal: *fulô, terém, saravá, parantá* (> *plantar*).

2. Todos os nomes, que no singular começarem pelas syllabas, ou letras abaxo, começaráõ no plurar em Ma [...].

E	Embe, Maembe	Pombos.
Y	Yala, Mayala	Machos.
Gi	Ngina, Magina	Nomes.
U	Uanga, Maüanga	Feitiços.
Co	Cota, Macota	Mais velhos.
Cu	Cunda, Macunda	Corcovas.
La	Lao, Malao	Riquezas.
Le	Leza, Maleza	Fraquezas. [...]

3. Os nomes, que começarem pelas letras abaxo, começaráõ no plurar em Gi [...].

G	Nganga, Padre	Ginganga, Padres.
U	Nvunda, Ginvunda	Brigas.
Z	Nzambi, Ginzambi	Deoses.
F	Fuba, Gifuba	Farinhas.
H	Hanga, Gihanga	Perdizes.
P	Pango, Gipango	Traças.
S	Sangi, Gisangi	Galinhas.
T	Tulo, Gitulo	Peitos.
X	Xitu, Gixitu	Carnes. (Dias, 1697: 4-5)

Como se vê, o quimbundo, ao contrário das línguas indo-europeias, como o PB, não usa terminações para marcar o plural, mas sim prefixos, partículas que são afixadas *antes* dos nomes. É possível, então, postular que os falantes de quimbundo, tendo de aprender português, identificaram nos *artigos* e outros determinantes dos nomes elementos semelhantes aos seus prefixos, de modo que não faziam a concordância dos nomes por meio de *flexões*. Disso teriam resultado concordâncias como *as casa, os menino, meus amigo, minhas terra, essas coisa*, tão características das variedades rurais e rurbanas brasileiras e também das variedades urbanas de prestígio, quando seus falantes se acham em situações de menor monitoramento, maior informalidade.

4. Eme, *Eu*, Eyè, *Tu*, Ae, *Elle*. Plur. Etu, *Nòs*, Enu, *Vòs*, Ao, *Elles*. [...] Naõ tem declinaçaõ, nem variedade de casos; como tem os pronomes Latinos, & servem de nominativos, & dos mais casos sem variedade dos ditos pronomes". (Dias, 1697: 8)

Em quimbundo, de fato, os pronomes pessoais são invariáveis, não mudam de forma segundo a função sintática (como em PB, *eu / me / mim*). É possível que venha daí a tendência que encontramos no PB rural e rurbano (e mesmo em determinados usos dos falantes urbanos de prestígio, [▶745]) de usar os pronomes do caso reto em todas as funções sintáticas: *"Eu como"*, *"Leva eu"*, *"Dá pra eu"*, *"Vem com eu"* etc.

5. Modo indicativo, tempo presente.

Sing. Nguizôla, *Eu amo.*
ūzôla, *tu amas.*
üzôla, *elle* ama.
Plur. Tuzòla, *nòs amamos.*
Muzòla, *vòs amais.*
Azòla, *eles amam.* (Dias, 1697: 13)

Tal como se dá com os nomes, também na conjugação verbal o quimbundo emprega prefixos e não terminações. Cada pessoa verbal tem seu prefixo próprio. No aprendizado irregular do português, os falantes de quimbundo teriam identificado nos pronomes pessoais do português os seus prefixos verbais, o que explica o paradigma de conjugação com presença obrigatória do pronome-sujeito e ausência de marcas flexionais de pessoa (com exceção da primeira) em tantas variedades do PB: *eu falo / tu fala / ele fala / nós fala / vocês fala / eles fala.*

O linguista alemão Eberhard Gärtner, especialista em português, enumera várias outras características do PB rural e rurbano que poderiam ser atribuídas ao contato de falantes das línguas bantas com o português colonial. Para reforçar sua tese, ele apresenta exemplos de fenômenos semelhantes no português angolano e moçambicano, usado como segunda língua por falantes de línguas bantas. Segundo Gärtner (2002: 294):

> Tanto as línguas africanas como as brasílicas pertencem a grupos tipológicos diferentes, que não conhecem o processo de flexão característico do português. Daí terá resultado que os falantes dessas línguas, os chamados aloglotas, obrigados a aprender o português num processo não guiado, simplificassem todos aqueles elementos do português cuja função não lhes fosse transparente do ponto de vista de sua própria língua.

Diz o autor que muitas das características próprias do PB se devem a "fenômenos que tiveram a sua origem no contato linguístico da época colonial, e que se expandiram da linguagem rural para a urbana inculta e daí para a fala culta" (2002: 297). Essas características seriam as seguintes:

- simplificação da flexão verbal;
- simplificação da flexão nominal (as línguas bantas não classificam as palavras segundo os gêneros);
- omissão (embora não sistemática) do artigo;
- omissão de preposições (inexistentes nas línguas bantas);
- extensão funcional da preposição *em* para indicar localização mas também direção (*Fui na venda*) — as línguas bantas não distinguem 'lugar *onde*' e 'lugar *aonde*': '*para* casa' e '*em* casa' se diz, indistintamente, '*ku* nzo' (quicongo), '*k*onjo' (umbundo) e '*ku* bata' (quimbundo);
- simplificação da flexão pronominal (como vimos acima);
- omissão do pronome pessoal objeto (sintaxe típica do PB, que nisso difere das demais línguas românicas: *Aluguei o filme, anda não vi [Ø], mas já emprestei [Ø] à Denise, que ficou de me devolver [Ø] no sábado.*);
- pronome sujeito em função de objeto (como vimos acima).

É impossível desconsiderar a profunda influência do elemento humano africano na cultura brasileira, da religiosidade à música, da culinária à indumentária e, evidentemente, na língua que se formou a partir desse intenso contato que alcançou, inclusive, as relações sexuais e a miscigenação.

5.9 A QUESTÃO DA "LÍNGUA BRASILEIRA"

Em seguida à independência, a pequena elite intelectual brasileira passará a se divertir com a questão da "língua brasileira". Esse debate se iniciou com o Visconde de Pedra Branca já em 1824-26. A polêmica se arrastará por longo tempo, e não se pode dizer que tenha terminado, muito embora se processe hoje em dia com base em teorias científicas mais consistentes, ao contrário do que sucedeu ao longo do século XIX e boa parte do século XX, quando a discussão foi empreendida em grande parte por não-especialistas. O reconhecimento da especificidade do PB como língua distinta se afirma no próprio uso do nome português brasileiro, como se faz na pesquisa linguística em nosso país há bom tempo. Também se destaca, na análise contemporânea, uma abordagem eminentemente discursivo-ideológica do problema e/ou mais centrada nos aspectos históricos, políticos e socioeconômicos, numa nítida percepção de que não se trata de uma questão de ordem unicamente linguística.

Convém destacar que, em suas distintas fases, essas discussões se processaram todas dentro de um pequeno círculo de letrados — essencialmente masculino, branco e oligárquico —, de dimensões ínfimas em relação ao restante da população, constituída de mulheres, que tinham acesso nulo ou restrito a uma educação formal, de milhões de escravos negros destituídos de seu estatuto de seres humanos e dos numerosos grupos étnicos e estratos sociais desprestigiados que desde sempre têm sido a grande maioria do povo brasileiro. Assim, os apelos inflamados em favor da "língua brasileira" compartilhavam, no fundo, o mesmo espírito elitista e conservador de seus supostos adversários, os defensores intransigentes da identidade linguística de brasileiros e portugueses. Como bem analisa o historiador Sérgio Buarque de Holanda: "Tradicionalistas e iconoclastas movem-se, em realidade, na mesma órbita de ideias. Estes, não menos do que aqueles, mostram-se fiéis preservadores do legado colonial, e as diferenças que os separam entre si são unicamente de forma e superfície". Assim é que nunca se fará a defesa das características linguísticas das variedades "populares", usadas pela maioria não branca da população.

O que se reivindicava, na verdade, era a legitimação de um escasso número de aspectos, sobretudo morfossintáticos (além de alguns "exotismos" lexicais), que, embora divergentes da norma-padrão lusitana, compareciam na língua falada dos brasileiros das classes privilegiadas. Realizações fonéticas como "paia" para o que se escreve *palha*, ou "pranta" para o que se escreve *planta*, ou ainda regras de concordância verbal do tipo "nós faz" nunca foram objeto dessas reivindicações por serem identificadas, antes de tudo, com classes sociais extremamente desprestigiadas, e até hoje constituem as variantes sociolinguísticas mais estigmatizadas pelas classes sociais escolarizadas. Reconhecer direito de cidadania a tais formas equivaleria "a nivelar por baixo", muito embora diversos traços das variedades populares já fizessem parte dos usos habituais dos cidadãos urbanos letrados, como é o caso, por exemplo, da cons-

trução "para mim fazer", que ocorre cada vez mais frequentemente na fala das camadas urbanas escolarizadas, sobretudo na cidade e no estado de São Paulo. [▶498]

O caso de José de Alencar (1829-1877) é bastante exemplar. A língua que o mais importante escritor do período romântico defende é, no final das contas, tão "brasileira" quanto o "índio" que ele retrata em seus romances — um índio idealizado, de espírito nobre e puro, o protótipo do "bom selvagem". Basta compararmos esse "índio" idealizado no passado com a situação do negro na obra de Alencar, negro escravizado no cotidiano do autor, um cotidiano incômodo e refratário a idealizações. Não se estranha, então, que ele tenha escrito que "não se pretende que toda inovação seja boa: defende-se a ideia do progresso da língua, não o abuso que a acompanha". Afinal, Alencar foi deputado pelo Partido Conservador e inimigo da abolição da escravatura, como deixa claro em suas *Cartas a favor da escravidão*, que só foram publicadas em conjunto em 2009, além de terem sido excluídas dos volumes da obra completa do escritor.

A luta de Alencar e de outros que viriam depois dele pela validação de pelo menos algumas características da língua portuguesa aqui falada não teve repercussão no ensino. Ele permaneceu sendo um ensino de gramática explicitamente voltado para a "língua portuguesa", isto é, a norma literária lusitana. É por isso que me referi a essa polêmica como um divertimento da nossa elite intelectual. No círculo mais amplo das relações sociais, inclusive educacionais, nunca foi abalado o prestígio de fato da norma-padrão lusitanizante.

Passada sua fase romântica, o ímpeto de nacionalismo elitista no que diz respeito à língua acabou vencido pelo projeto europeizador maior da oligarquia, sobretudo com a ascensão intelectual do Parnasianismo, marcado pelo preciosismo sintático e lexical fortemente classicizante. Nessa escola literária, mais que nunca, qualquer elemento "popular" deveria ser evitado ao máximo. É notável a aparente "contradição" entre forma e fundo do citadíssimo soneto de Olavo Bilac (1865-1918) "À língua portuguesa", arquétipo do parnasiano, em que ao lado de expressões como "tuba de alto clangor" e "trom e silvo da procela", se lê que foi nessa língua que "da voz materna ouvi: 'meu filho'!" Nenhuma mãe brasileira, de nenhuma classe social, jamais falaria ao filho naquela língua contorcida e cinzelada a golpes de latinismos.

O apogeu dessa fase de culto a uma norma lusitanizante se concretiza, fisicamente inclusive, na criação da Academia Brasileira de Letras (ABL) em 1896, que teve entre seus fundadores vários dos mais destacados representantes da escola parnasiana entre nós. Como toda instituição de seu gênero, imitação do modelo aristocrático-feudal da Academia Francesa, a ABL perpetua até hoje aquele espírito de "defesa" da "língua" (isto é, da norma-padrão tradicional lusitanizada) contra toda "corrupção" provocada pelo (ab)uso do idioma por parte do "vulgo". Ela quer preservar *ad immortalitatem* uma atitude entranhadamente elitista diante dos fatos linguísticos,

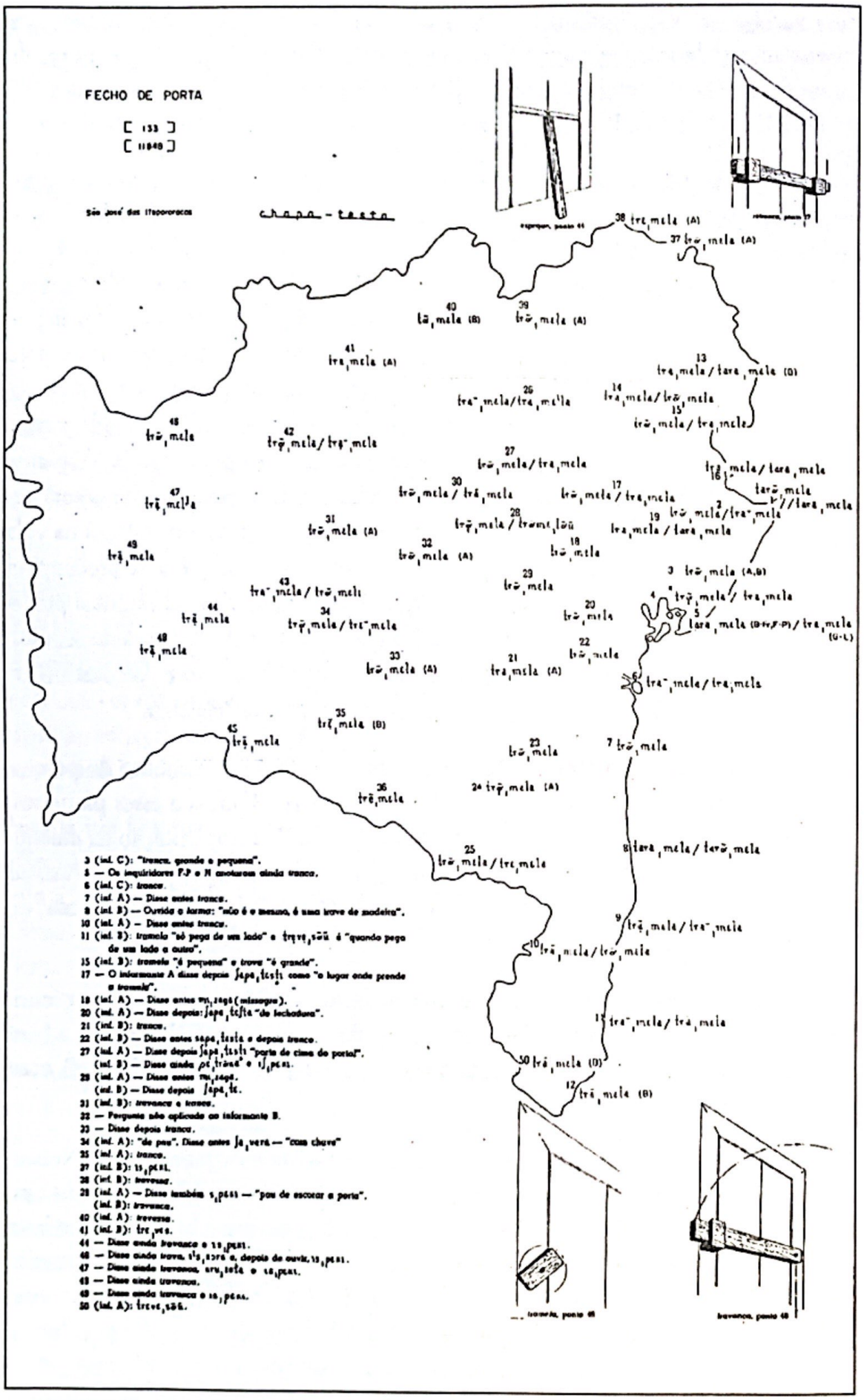
uma vez que seus quarenta membros (quase todos ignorantes dos desdobramentos da pesquisa e da teorização científica acerca da língua) se consideram capazes de tomar decisões e decretar escolhas em detrimento das opções de uso dos demais 190 milhões de falantes da língua majoritária do Brasil, inclusive dos mais letrados.

No início do século XX, a polêmica em torno da "língua brasileira", se reacende, novamente encampada por um movimento estético-literário, o Modernismo. Exemplar, nesse período, é o projeto (nunca levado a cabo) de Mário de Andrade (1893-1945) de sistematizar uma "Gramatiquinha da língua brasileira". Tal como no período romântico, a defesa do "brasileiro" será feita em grande parte por pessoas desvinculadas da pesquisa linguística sistemática e da prática docente e bem mais interessadas em questões de vanguardismo estético, de definição de uma identidade nacional, de aproveitamento do folclore (sobretudo o folclore de raiz indígena e africana) etc. Quando muito, se valerão das conclusões das pesquisas empreendidas pelos pioneiros da incipiente dialetologia brasileira. A defesa da "língua brasileira", mais uma vez, permanecerá circunscrita às esferas da intelectualidade e sem praticamente nenhum influxo na vida dos cidadãos comuns e dos poucos que tinham acesso à escola, que só se preocupava em inculcar uma norma linguística conservadora, ensinada com os mesmos procedimentos filológicos e com a mesma didática com que se ensinou, durante séculos, a língua latina no Ocidente: o português era analisado, dissecado e "reconstituído" como se fosse uma língua morta, preservada apenas em textos "clássicos".

Na primeira metade do século XX, ocorre um importante movimento de pesquisa das diferentes variedades linguísticas faladas no Brasil. Graças a esses pioneiros, temos registros de alguns importantes falares regionais. É a implantação da *dialetologia* entre nós, com obras importanes como *O dialeto caipira*, de Amadeu Amaral (1920); *O linguajar carioca*, de Antenor Nascentes (1922) e *A língua do Nordeste*, de Mário Marroquim (1945).

Esses trabalhos logo serão seguidos de muitos outros, empreendidos por pessoas interessadas em conhecer a realidade linguística do Brasil, mas ainda sem uma base teórica e metodológica rigorosa, que só viria a surgir depois da década de 1970, com o advento da sociolinguística.

Na década de 1960, graças sobretudo aos empreendimentos pioneiros de Nelson Rossi, tem início no Brasil uma importante corrente de pesquisas geolinguísticas, isto é, investigações sobre a diversidade linguística do português brasileiro com base em levantamentos de usos nos diferentes pontos do território nacional. O primeiro trabalho de destaque é, sem dúvida, o monumental *Atlas prévio dos falares baianos*, elaborado por uma equipe de linguistas baianas e baianos, coordenados por Rossi e publicado em 1963. A esse projeto inicial se seguiram vários que confluíram para o grande projeto do *Atlas linguístico do Brasil* (ALiB), ainda em andamento, que conta com a participação de estudiosos das diversas regiões do país (Cardoso, 2010).



[Mapa 5.11 - *Atlas Prévio dos Falares Baianos* — Reprodução da Carta 152/ Fecho da Porta.]

É precisamente no início dos anos 1970 que se lança o Projeto NURC (Norma Urbana Culta), um divisor de águas nos estudos sobre o PB. Com base nas centenas de horas de gravação da fala de pessoas nascidas e criadas em zona urbana e com nível superior completo, um amplo número de pesquisadores pôde se dedicar à investigação do PB sem precisar recorrer à própria intuição ou a exemplos inventados: existia agora um grande banco de dados empíricos, testemunhos audíveis da fala urbana de prestígio. A principal descoberta do NURC foi a tremenda distância entre a língua realmente empregada por essas camadas sociais mais letradas e a concepção de "língua certa" prescrita pela tradição normativa. Assim, além das diferenças entre a norma popular e a norma culta, conhecidas desde o fim do período colonial, foi possível demonstrar que o PB urbano culto também era radicalmente distinto da norma-padrão clássica, inspirada nos usos literários de autores consagrados do Romantismo português. Os frutos do NURC são colhidos até hoje, pois a partir desse mesmo projeto foi possível elaborar um outro, igualmente importante: a Gramática do Português Falado. Uma vez concluída a publicação das obras previstas, o PB será a primeira grande língua da família românica a ter suas variedades urbanas de prestígio completamente descritas e analisadas.

Também a partir do final dos anos 1970 têm início as pesquisas baseadas na metodologia desenvolvida nos Estados Unidos por William Labov, considerado o nome mais importante da sociolinguística variacionista. Um número cada vez maior de variedades linguísticas — urbanas, rurais, rurbanas, de falantes menos escolarizados, mais escolarizados, de comunidades isoladas, de homens, de mulheres, de jovens, de idosos, de imigrantes etc. — vai sendo documentado e estudado. Hoje não é mais possível dizer, como era possível cinquenta anos atrás, que a realidade linguística brasileira é pouco conhecida. Esse enorme volume de conhecimento sobre o PB tem permitido uma análise crítica das práticas tradicionais de educação em língua materna. A pesquisadora Stella Maris Bortoni-Ricardo abriu uma nova perspectiva de investigação e de aplicação prática, inaugurando o campo da *sociolinguística educacional* no Brasil.

Todo esse grande movimento científico tem resultado na tomada de consciência cada vez mais aguda de que o PB é uma língua diferente do português europeu, com o qual tem fortes laços de parentesco, mas do qual já se afastou, tanto em suas variedades mais urbanas e de prestígio quanto em suas variedades "populares" e estigmatizadas. Mesmo os linguistas que operam com um conceito de língua restrito ao sistema e às suas estruturas, como os da escola gerativista, reconhecem a necessidade de considerar o PB e o português europeu como duas línguas distintas, já que têm duas *gramáticas* distintas — *gramática* entendida aqui como a capacidade que todo falante tem de produzir sentenças bem formadas em sua língua materna. Se levarmos o conceito de língua para um plano mais amplo que inclua as questões sociais, culturais, históricas etc., é simplesmente imperioso reconhecer o PB como uma língua plena. Não importa se chamamos a língua de "português", "português brasileiro" ou "brasileiro". O que importa — para os profissionais engajados no estudo e no ensino da língua — é abandonar a ineficaz comparação entre o PB e o português europeu e, ainda

mais ineficaz, entre o PB e a norma-padrão clássica. Não só ineficaz como também perniciosa, pois dessa comparação o PB sempre saiu "perdendo", porque visto como "defeituoso", "falho" ou "estropiado".

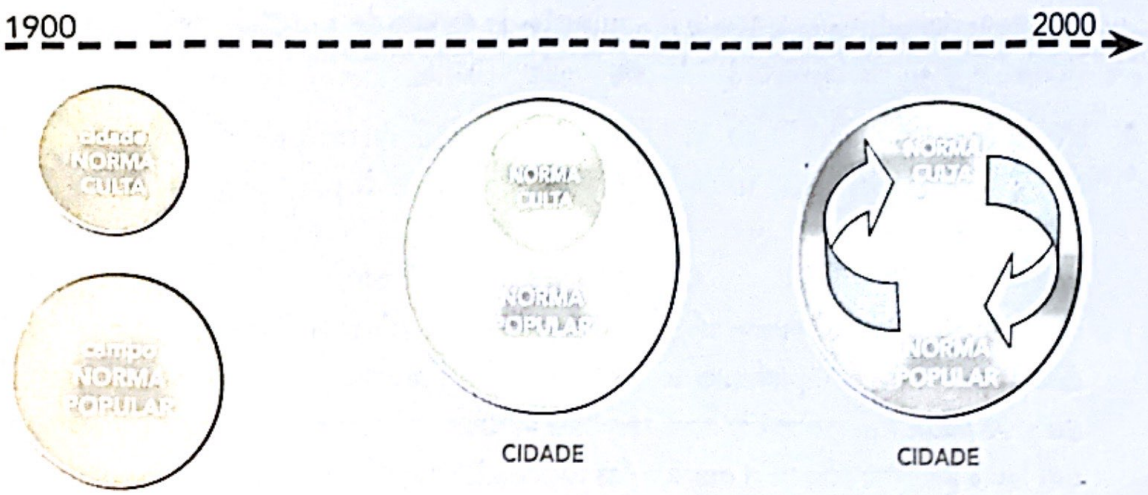
5.10 A REALIDADE SOCIOLINGUÍSTICA DO PB: DA POLARIZAÇÃO À NEBULOSA

Os fatos históricos abordados acima explicam a situação de polarização sociolinguística do PB (termo sugerido por Lucchesi, 2002): de um lado, as variedades urbanas de prestígio (agrupadas tradicionalmente sob o rótulo de *norma culta*), herdeiras das variedades de português europeu trazidas para cá pelos brancos e transmitidas por seus descendentes de geração em geração, e empregadas por uma reduzida porção da população total; do outro, as variedades rurais e rurbanas (agrupadas sob o rótulo de *norma popular*), estigmatizadas pelos usuários das variedades urbanas de prestígio, empregadas pela maioria absoluta dos brasileiros.

É a formação social do Brasil que explica essa polarização, pois para cada um desses dois conjuntos de variedades é possível retraçar uma história diferente: as variedades que compõem a *norma popular* são o prosseguimento do português adquirido de forma irregular pelos escravos, mestiços e seus descendentes, e reestruturado em seus grandes traços morfossintáticos e possivelmente fonológicos; as variedades que compõem a *norma culta* são a continuação do português empregado pelas elites brancas urbanas, sempre tendo como modelo a norma de prestígio de Lisboa.

Essa separação radical entre a norma culta e a norma popular decerto predominou por um longo período da nossa história. A partir da segunda metade do século XX, porém, tem início o intenso processo de urbanização da população brasileira. Se em 1960, segundo dados do IBGE, somente 45% da população brasileira vivia nas cidades, quarenta anos depois, o censo de 2000 revelou que 80% dos brasileiros viviam em zonas urbanas.

Essa acelerada e caótica urbanização causou o inchaço populacional das nossas cidades, especialmente das capitais e polos econômicos importantes. As periferias e favelas das grandes cidades se tornaram o ambiente mais importante de cultivo e preservação da *norma popular* brasileira. A polarização que antes podia ser traçada em termos de *cidade* vs. *campo* agora se transfere para as relações entre as camadas sociais presentes nas áreas metropolitanas.



As complexas relações e interações sociais no ambiente urbano vão aos poucos transformando o panorama sociolinguístico brasileiro. É certo que ainda hoje é possível reconhecer uma fronteira linguística entre as variedades urbanas de prestígio e as variedades urbanas e rurbanas estigmatizadas. No entanto, essa fronteira é muito mais "líquida" e difusa do que antigamente, a ponto de hoje ser, de fato, nebulosa. Isso tem permitido que traços linguísticos que até pouco tempo eram características exclusivas da norma popular avancem em direção às camadas economicamente superiores da população, e que formas linguísticas próprias dessas camadas "cultas" passem a ser empregadas também pelos cidadãos de menor poder aquisitivo e de acesso mais restrito à educação formal.

Vários fenômenos têm contribuído para esse nivelamento sociolinguístico da sociedade brasileira contemporânea:

- a difusão maciça das tecnologias de comunicação, com destaque especial para a televisão, que chega a mais de 90% dos lares brasileiros; como se sabe, a televisão foi um meio empregado especialmente pela ditadura militar (1964-1984) para promover a integração da população, reduzir as distâncias, difundir a ideologia oficial e incentivar o consumo. Com isso, a programação televisiva, produzida integralmente nos grandes centros urbanos, se tornou um agente de difusão das normas urbanas de prestígio, principalmente das variedades carioca e, mais tarde, paulistana. É na televisão, especialmente nos telejornais, que podemos localizar uma espécie de pronúncia padrão brasileira [▸339]. O mesmo vale para o rádio;
- a massificação do ensino público: a urbanização repentina provocou a chamada "democratização" do ensino, que representou de fato uma expansão *quantitativa* do universo escolar e uma igual deterioração *qualitativa* da educação e das condições de ensino-aprendizagem, acompanhada da progressiva depreciação da profissão docente. Essa educação pública, ainda que reconhecidamente ruim, permitiu o acesso de um número cada vez maior de pessoas às instâncias e aos agentes de letramento. Nesse processo, o contato com as normas urbanas de prestígio, representadas principalmente nos textos destinados à leitura presentes

nos materiais didáticos, levou à aquisição de traços de prestígio por parte das camadas mais desfavorecidas da população urbana;

- a deterioração da profissão docente fez ela ser abandonada pelas camadas médias e altas da sociedade, usuárias das normas urbanas de prestígio. Aos poucos, a docência passou a ser exercida majoritariamente por professoras e professores vindos das classes C, D e E. Com isso, o projeto ideológico da escola (suposto mas, de fato, jamais cumprido) de transmitir e inculcar o "bom português" se dissolveu quase completamente, uma vez que a antiga separação social entre *docente falante de norma culta* e *aprendiz falante de norma popular* desapareceu em favor de uma presença maciça das variedades populares no ambiente escolar, usadas pelos aprendizes, pelos docentes e pelos demais agentes educacionais;

- a redemocratização da sociedade, a partir dos anos 1980, e a progressiva participação do Brasil nas engrenagens da assim chamada "globalização". Aos poucos se instaura no Brasil, em grupos sociais cada vez mais amplos, a tendência ideológica rotulada de "pós-moderna", que enfatiza o papel da linguagem, das relações de poder e das motivações individuais, combatendo em particular as classificações e divisões estanques e rígidas como masculino/feminino, heterossexual/homossexual, negro/branco, certo/errado etc. É claro que isso não elimina o forte preconceito linguístico que sempre caracterizou as relações sociais no Brasil, mas pelo menos tornou ele evidente e suscitou críticas e denúncias dessa discriminação social por meio da linguagem;

- mais recentemente, a ascensão da chamada "nova classe média", composta pelos cidadãos das classes C e D que vêm progressivamente se inserindo no mercado de consumo, alcançando a educação de nível superior, conquistando acesso a bens e serviços até então reservados às camadas privilegiadas (aquisição da casa própria, turismo internacional, viagens de avião, telefonia móvel e fixa, televisão paga, automóvel novo, internet etc.). Essa nova classe média sem dúvida vai levar para o topo da pirâmide social brasileira as formas linguísticas "populares", fazendo emergir aos poucos uma *nova norma urbana culta*, bastante diferente daquela que os pesquisadores brasileiros registraram no início dos anos 1970 para compor os dados do Projeto NURC.

A eleição, em 2002, de Luiz Inácio Lula da Silva à presidência da República simboliza claramente essa transição de uma *polarização* sociolinguística rumo a uma *nebulosa* sociolinguística. É preciso que fique claro, no entanto, o caráter *simbólico* dessa eleição, já que as relações de poder na sociedade brasileira não foram estruturalmente modificadas e sim, bem mais, reconfiguradas, de modo a se formar um compromisso entre os novos donos do poder e as elites socioeconômicas e socioculturais dominantes. Com isso, muitos avanços sociais importantes foram obtidos, como uma significativa redução da pobreza, sem que os privilégios das classes dominantes fossem significativamente ameaçados.

Ainda mais recentemente (2010), a eleição de Dilma Rousseff à presidência da República também traz uma forte carga simbólica, pois se trata da primeira mulher a assumir esse cargo. É de esperar que, pela própria história de vida de Rousseff, se empreenda no atual governo um projeto social mais progressista e menos comprometido com os interesses oligárquicos.

5.11 A REAÇÃO À DEMOCRATIZAÇÃO DAS RELAÇÕES SOCIOLINGUÍSTICAS

Numa sociedade como a brasileira, marcada desde suas origens por uma divisão social profunda entre uma minoria privilegiada e uma imensa maioria marginalizada, a língua sempre foi um símbolo importante para as elites dominantes. "Saber português" é uma condição tida como *sine qua non* para que alguém se incorpore ao círculo dos que podem – é o melhor exemplo daquilo que Pierre Bourdieu chama de distinção. No entanto, o suposto conhecimento do "português correto" é uma mera fachada para justificar o emprego de outros instrumentos, mais complexos e sofisticados, destinados a preservar a desigualdade social. Afinal, num país tão racista, se a pessoa for negra ou índia, por exemplo, por melhor que conheça a norma-padrão tradicional, já estará excluída do acesso ao círculo dos privilegiados pela simples cor de sua pele. O mesmo se diga com relação a nordestinos no Sudeste, a homossexuais ou a mulheres, a cegos, surdos, deficientes físicos, em muitos âmbitos da vida social, econômica e política.

A difusão do preconceito linguístico não é novidade na história do Brasil. Pelo contrário, é uma das marcas da nossa formação histórica: desde o Diretório dos Índios, decretado pelo Marquês de Pombal no século XVIII para proibir o ensino das línguas indígenas e impor o português, passando pelo extermínio sistemático de centenas de povos indígenas e suas línguas, pela distribuição dos escravos africanos em lotes contendo indivíduos que não falavam as mesmas línguas, chegando ao século XX com o "crime idiomático" definido pela ditadura Vargas para coibir o uso do alemão, do italiano e outras línguas de imigração durante a II Guerra Mundial

Entretanto, em sua versão mais recente, essa difusão parece ser motivada por uma reação de alguns setores da elite mais letrada contra o nivelamento sociolinguístico que vem sendo detectado há pelo menos cinquenta anos na sociedade brasileira, devido sobretudo ao processo de intensa urbanização da nossa população, que trouxe para as grandes cidades as variedades linguísticas mais estigmatizadas, características da zona rural e, hoje, do cinturão rurbano que rodeia as nossas metrópoles.

O discurso em "defesa" da "língua portuguesa" supostamente ameaçada (por seus próprios falantes nativos...) começou a se tornar mais audível e visível nas manifes-

tações de purismo linguístico dos grandes meios de comunicação. A partir dos anos 1990, esse fenômeno — que gosto de chamar de *comandos paragramaticais* — ganha amplo espaço na televisão, no rádio, nos jornais, nas revistas e na internet. Ressurgem, com roupagem tecnológica avançada, os antigos "consultórios gramaticais" da imprensa da primeira metade do século XX, quando imperavam figuras como o filólogo português Cândido de Figueiredo, nada cândido em suas diatribes reunidas sob o título parnasiano O que se não deve dizer, com apossínclise e tudo, e o gramático brasileiro Napoleão Mendes de Almeida, autor de um Dicionário de questões vernáculas onde o preconceito social é explícito e assumido contra "cozinheiras, engraxates, trombadinhas e infelizes caipiras", conforme suas próprias palavras.

As grandes empresas de comunicação parecem querer assumir o lugar de depositárias das "belas letras" e do "bom português" que a literatura moderna e contemporânea foi abandonando aos poucos, graças às novas tendências estéticas de experimentação com a linguagem, de valorização e estilização das variedades linguísticas desprestigiadas, de veiculação de crenças e ideologias antielitistas, de concessão de voz literária às camadas marginalizadas da sociedade, de hibridização intensa dos gêneros literários tradicionais etc.

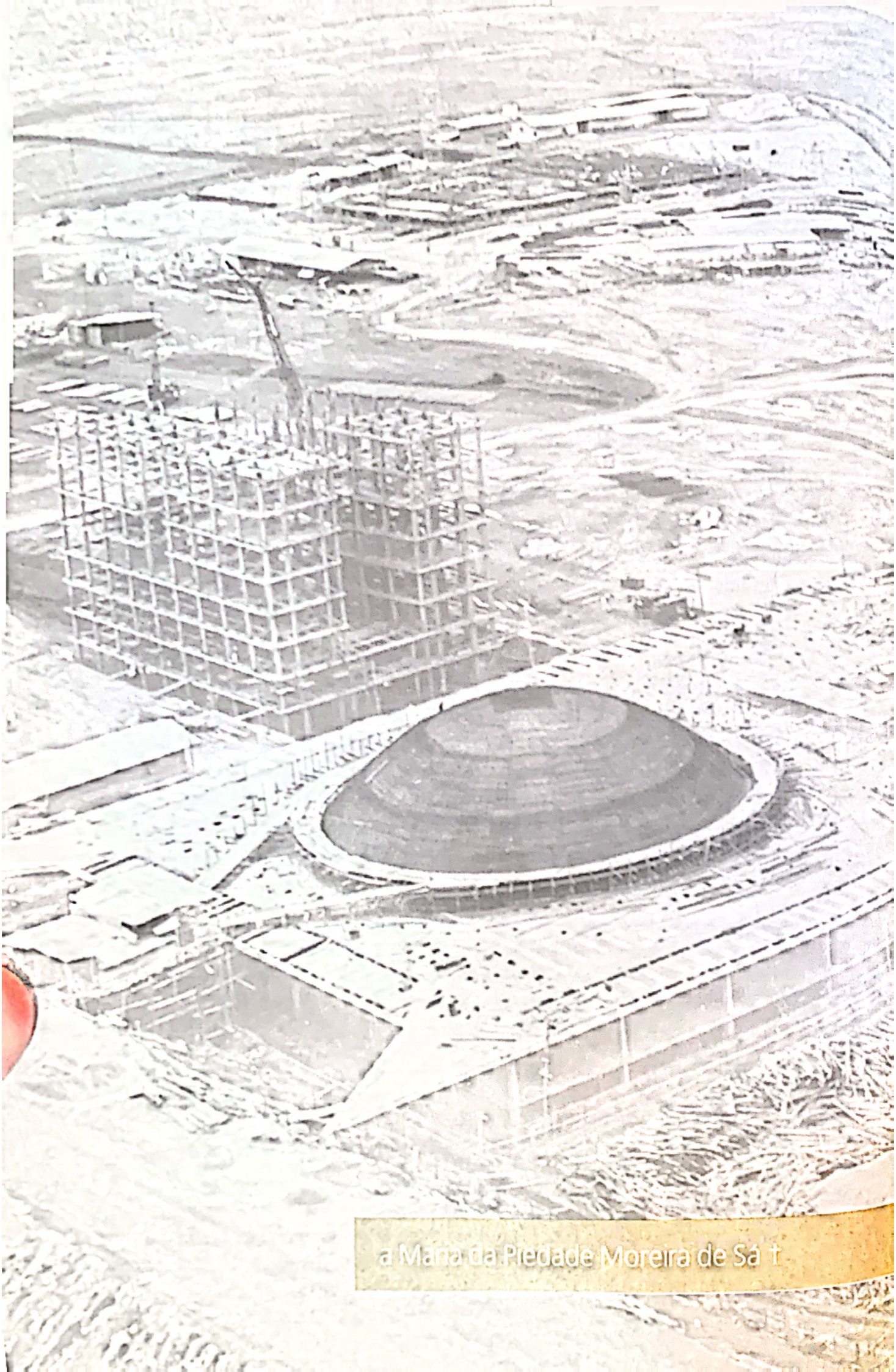
Essas manifestações reacionárias do purismo linguístico nos meios de comunicação se chocam violentamente contra as diretrizes oficiais de ensino de língua, que – no Brasil e no mundo – procuram incorporar e aplicar os avanços das pesquisas da linguística teórica e aplicada. Parâmetros curriculares nacionais, programas de formação continuada de docentes, programas de incentivo à leitura, de distribuição de material didático, entre tantos outros levados a efeito pelas instâncias superiores da educação brasileira, negam peremptoriamente a validade do discurso purista diante do conhecimento acumulado pelas ciências da linguagem acerca da realidade do português brasileiro:

> A imagem de uma língua única, mais próxima da modalidade escrita da linguagem, subjacente às prescrições normativas da gramática escolar, dos manuais e mesmo dos programas de difusão da mídia sobre "o que se deve e o que não se deve falar e escrever", não se sustenta na análise empírica dos usos da língua. (*Parâmetros Curriculares Nacionais*. Língua Portuguesa. 3º e 4º ciclos do Ensino Fundamental. Brasília: MEC, 1998, p. 29)

Infelizmente, porém, essas vozes puristas têm grande audiência entre nós. Isso se deve, entre outras coisas, à transformação da "língua certa" em objeto de desejo precisamente das emergentes camadas C, D e E, que reconhecem em seus modos de falar grandes diferenças com relação aos usos considerados "cultos" e tentam conquistar essa língua idealizada com vistas a uma suposta ascensão social que esse conhecimento permitiria. Essa demanda é atendida pelo mercado com a transformação dessa "língua" num bem de consumo supostamente acessível a todos e disponível sob as mais diferentes embalagens e modelos (programas de televisão e de rádio, colunas de jornal e de revista, programas para computador, CD-roms, livros,

revistas, fascículos, *sites* na internet, cursos, tira-dúvidas por telefone, manuais de redação das grandes empresas jornalísticas etc.). Ora, a estratégia publicitária clássica para a venda de qualquer produto é convencer o potencial consumidor da necessidade de preencher alguma carência essencial. Assim, num país em que o acesso à boa educação sempre foi restrito e em que se cristalizou na mentalidade comum o mito de que "brasileiro não sabe português", nada é mais fácil do que conquistar essa clientela ávida por uma "língua" boa, segura e com selo de qualidade conferido por autoproclamados especialistas na matéria.

A ignorância e a arrogância dos meios de comunicação no tratamento de questões de linguagem têm sido amplamente examinadas e discutidas por linguistas brasileiros. As pessoas interessadas podem ler com grande proveito os livros de Maria Marta Pereira Scherre (2005) e Sírio Possenti (2009).

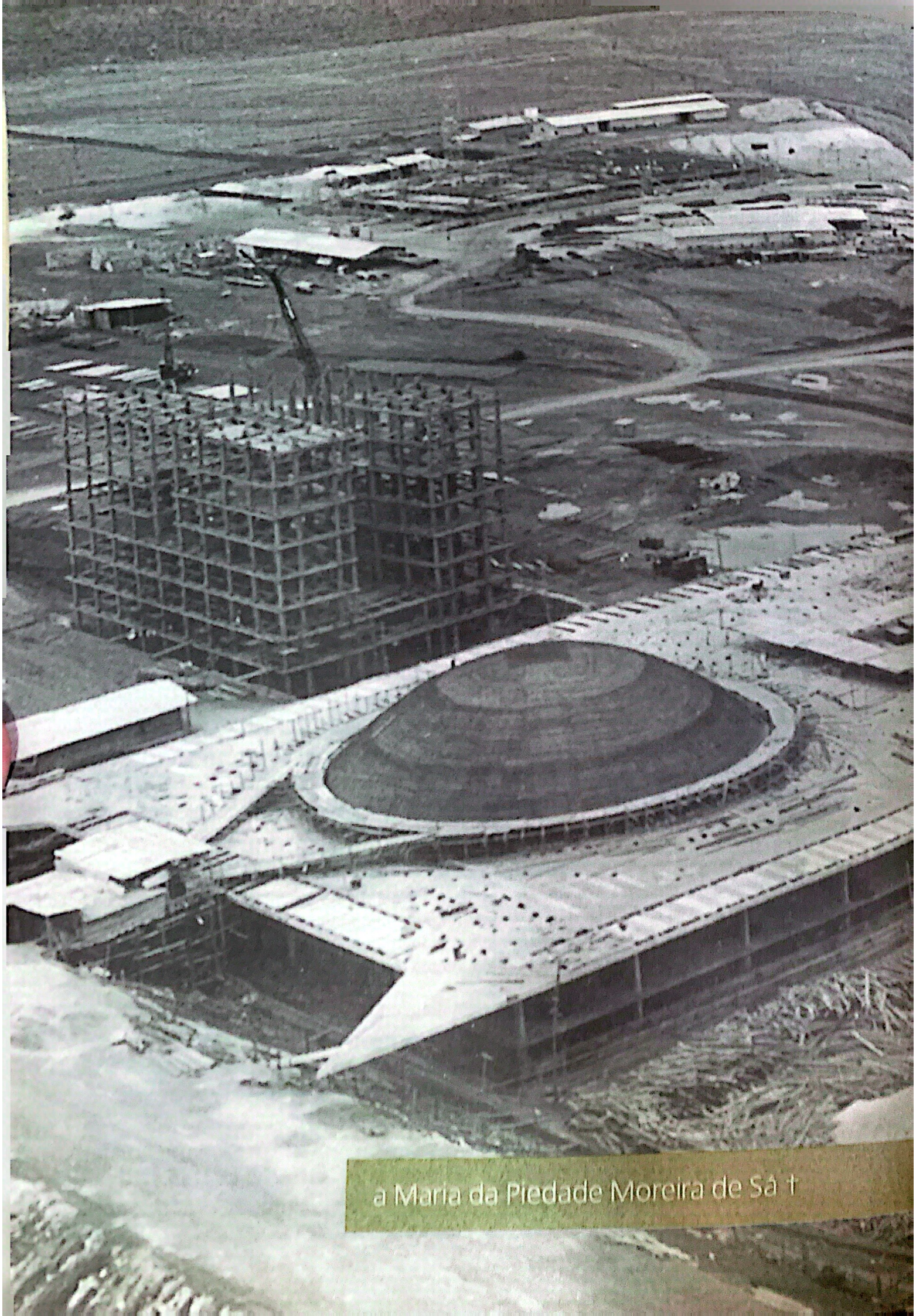


a Maria da Piedade Moreira de Sá †

6

raízes desterradas

formação do léxico português



a Maria da Piedade Moreira de Sá †

6

raízes desterradas

formação do léxico portugalego

Muitas pessoas exibem uma grande curiosidade quanto à origem das palavras, principalmente dos nomes próprios e, é claro, de seu próprio nome. O estudo dessas origens se chama etimologia e o termo do qual uma palavra se originou é chamado de étimo.

A palavra *étimo* tem, ela mesma, um étimo interessante. Vem do grego *étymon*, que significava "sentido verdadeiro de uma

palavra" e está, uma vez mais, relacionada com as investigações dos primeiros filósofos gregos que se debruçaram sobre a linguagem humana. Para esses filósofos, especialmente Platão, as palavras vão com o tempo sofrendo "corrupções" de sentido, à medida que são usadas pelas pessoas, afastando-se, desse modo, de um suposto sentido "original" que teriam quando foram "criadas" pelo Demiurgo, o artesão divino ou o princípio organizador do universo. Por isso, buscar o *étimo* de uma palavra seria encontrar o sentido original, primeiro, que ela teria. Na terminologia linguística, porém, essa ideia de "verdade" foi abandonada e hoje a etimologia é simplesmente o estudo da origem de uma palavra.

Uma coisa, no entanto, é preciso ficar muito clara. A etimologia é uma disciplina extremamente complexa, que exige, antes de mais nada, um profundo conhecimento da história da língua, no nosso caso o PB, um domínio amplo do latim, principalmente, e das línguas derivadas dele (para efeitos de comparação), bem como do grego e, se possível, do indo-europeu. Além disso, o estudioso deve recorrer às contribuições da história, da geografia, da arqueologia, da antropologia e ser capaz de consultar textos antigos e levantar as palavras e seus significados ao longo do tempo. Por isso, é sempre bom desconfiar de etimologias oferecidas por leigos sem formação adequada para tratar do assunto.

Não acredite, por exemplo, se alguém lhe disser (ou se você ler) que a palavra *aluno*, do latim *alumnu-*, é formada de *a-*, com sentido privativo ("sem") e de *lumen* ("luz") porque essa é uma lorota do tamanho de um bonde. A palavra *alumnu-* deriva do verbo latino *alo* (infinitivo: *álere*) com o sentido de "fazer aumentar, crescer, desenvolver; nutrir, alimentar, criar, sustentar, produzir, fortalecer" etc. O *aluno*, portanto, é aquele que está sendo nutrido, alimentado e, por metáfora, nutrido de conhecimento, alimentado de saber. A própria palavra *alimento* traz esse verbo em sua composição: *ali.mentum*. O particípio passado desse verbo é *altu-*, que nos deu o adjetivo *alto*, ou seja, originalmente, "crescido, desenvolvido", e toda a enorme quantidade de palavras derivadas dele. Também não acredite nas pessoas que tentam relacionar *adolescente* com *doente* (do latim *dolente-*), mais uma etimologia tirada do bolso do colete. *Adolescente*, por menos que pareça, se origina igualmente do verbo *alo*. Por efeito da apofonia, que vamos examinar logo adiante [▶270] a raiz *al-*, ao receber o prefixo *ad-*, se transforma em *ul-*: *ad.ul.escente*. O adolescente é aquele que está em fase de crescimento, assim como o *ad.ul.tus*, o adulto, é aquele que já concluiu sua formação.

A etimologia, portanto, é trabalho para gente bem preparada, que nunca vai dizer, como se ouve tão frequentemente, que *enfezado* tem a ver com "fezes", já que essa palavra, como nos explica o dicionário *Houaiss*, vem do "latim *infensátum* > *infensado*, 'encarniçado contra, hostil', part. pas. do v. lat. *infenso,as,ávi,átum,áre* 'encarniçar-se contra, ser hostil a', numa evolução semelhante à de *defesa/devesa* [...]; admitir essa nova base etimológica implica acolher *enfesado* como a grafia que se justifica, em lugar da grafia *enfezado*, historicamente equivocada".

Outro grave problema referente à etimologia é a atitude de muitos puristas de exigir que, no PB contemporâneo, as palavras sejam usadas de determinada maneira somente porque, *em latim clássico*, elas tinham tal significado ou tal uso. Por exemplo, não aceitam o uso do pronome *lhe* como objeto direto ("Eu não *lhe* conheço") porque, *em latim*, esse pronome tinha função apenas de objeto *indireto*. Esse argumento não faz o menor sentido: primeiro, porque não falamos latim, falamos uma língua derivada dele, sim, mas que passou por diversas e profundas transformações nos últimos mil e quinhentos anos; segundo, porque fazemos muitos usos admitidos nas gramáticas normativas e que não existiam em latim (por exemplo, o uso de pronome reflexivo no verbo *suicidar-se*). Os puristas se caracterizam por uma tentativa (sempre frustrada) de reprimir a mudança linguística e de considerá-la (como se fazia antes de Cristo!) uma "corrupção" ou uma "ruína" da língua [▶101].

O léxico do PB, como o de praticamente todas as línguas do mundo, é formado de diferentes "correntes migratórias" que, ao longo dos séculos, foram deixando suas contribuições. Por ser uma língua *românica*, o PB tem uma herança lexical *latina* importantíssima, responsável pela absoluta maioria das palavras que usamos.

6.1 Herança pré-romana

Antes de tratarmos da herança latina, porém, vamos ver de que forma os *substratos* [▶134] deixaram seus vestígios no nosso léxico. Quanto aos *adstratos* e *superestratos*, já vimos algumas de suas contribuições em [▶217]. O afluxo de palavras de origem *tupi* e *banta* no PB também foi abordado em [▶231/238].

Como examinamos em [▶134], a Península Ibérica, quando foi invadida pelos romanos no século II a.C., era habitada por diferentes povos, falantes de diferentes línguas, algumas remotamente aparentadas entre si (como as célticas, o lusitano e o grego: indo-europeias) e outras de famílias sem nenhum parentesco (ibérica, fenícia, tartéssia), para não mencionar algumas dúvidas sobre a presença de outros povos, como os **ambronos**. Devido à antiguidade dessas civilizações e à extrema escassez de fontes confiáveis, é dificílimo, quando não impossível, atribuir determinado vocábulo a essa ou àquela língua pré-romana.

Muitos **topônimos** (nomes de lugar) e **potamônimos** (nomes de rios) da Península Ibérica têm origem nessas línguas muito remotas. Em Portugal, por exemplo, os nomes de lugar Armona, Braga, Bragança, Cambra, Coimbra, Évora, Gaia, Juromenha, Lagos, Lamego, Langroiva, Lisboa, Pedrógão, Pióbão, Portugal, Terena são de origem **pré-romana**. O nome Lisboa parece vir do fenício *Allis Ubbo*, "porto seguro". Entre os rios portugueses, muitos têm nomes igualmente pré-romanos, ainda que pareçam palavras normais da língua: Ardila, Arnoia, Ave, Cávado, Coa, Dão,

Douro, Erges, Gilão, Homem, Inha, Leça, Lena, Lima, Mira, Mondego, Neiva, Paiva, Rabagão, Sabor, Sisandro, Sor, Tâmega, Tua, Ul, Vez, Vouga, Xévora.

Essa é uma situação muito comum na história das conquistas de um povo por outro: mesmo impondo sua língua, o conquistador acaba preservando os nomes próprios (cidade, rio, montanha etc.) com que o povo conquistado, ancestralmente, designava esses lugares. É o que ocorre no Brasil, onde temos uma quantidade incontável de topônimos de origem indígena, principalmente tupi (Pará, Paraíba, Paraná, Paranaguá, Paranapanema, Paranapiacaba, para citar alguns poucos que têm o elemento *pará*, *paraná*, 'mar, rio').

Vamos ver a seguir algumas conhecidas palavras de nossa língua que têm provável origem pré-romana:

PALAVRA	ORIGEM
ABÓBORA	desconhecida, mas certamente hispânica
ARROIO	provavelmente basca
BALSA	desconhecida, mas certamente hispânica
BARATO	pré-romana, talvez celta; só ocorre em portugalego e espanhol
BARRA	do gaulês **barro*, 'extremidade'
BARRACO	do assírio *parakku*, levado pelos fenícios à Península Ibérica
BARRANCO	provavelmente da mesma raiz de *barro*
BARRICA	do gaulês **barrika*, mesma origem de *barril* e *barriga*
BARRIGA	de *barrica*, com abrandamento do [k] intervocálico em [g]
BARRO	talvez celta, com muitas derivações nas línguas ibéricas
BEIÇO	origem incerta, talvez celta; só ocorre em portugalego e espanhol
BEZERRO	ibérica, de **ibikirru*, 'cabrito', mesma origem de *íbex*, 'cabrito-montês'
BREJO	obscura, talvez celta
BRIO	do celta **brigos*, 'força, coragem'
BRUXA	obscura, pré-romana; só ocorre em portugalego e espanhol
BURACO	pré-romana indo-europeia, aparentada talvez ao verbo latino que nos deu *furar*
CAMA	ibérica
CAMINHO	celta
CANTO	'pedra de grande tamanho', provavelmente celta; só ocorre em portugalego, espanhol e catalão; origem de *cantaria*, *canteiro*, *alcantilado* etc.
CARRASCO	de uma raiz ibérica **karr-*, 'carvalho'; no sentido de 'algoz', deriva do nome de Belchior Nunes Carrasco, algoz que teria vivido em Lisboa no século XV
CARVALHO	da raiz ibérica **karr-*
CHARNECA	do ibero-basco *sarna*, com pronúncia *txarna*
COELHO	do latin *cun.iculus*, de uma provável raiz celta **kun*, 'cão', (cp. grego *kyn-*, 'cão')

ESQUERDO	de alguma língua pré-romana que deu origem ao basco *ezkerr*
GARRA	provavelmente do celta **garra*, 'mão'
GORDO	obscura, provavelmente hispânica
LAJE	talvez celta
LAPA	pré-céltico, só ocorre em portugalego
LÉGUA	do gaulês
LOUSA	celta, 'pedra achatada'
MANTEIGA	obscura, pré-romana, só ocorre nas línguas da Península Ibérica
MARANHA	pré-romana; daí *emaranhar* etc.; só ocorre em portugalego e espanhol
MARRA	pré-romana, talvez celta
MOÇO	pré-romana, talvez ibérica; só ocorre em portugalego e espanhol
MODORRA	provavelmente ibérico, cp. com o basco *mutur*, 'enjoado, incomodado'
MOITA	obscura, pré-romana
MORRINHA	obscura; em galego é a palavra usada para o que chamamos de *saudade*
PEÇA	celta **pettia*, 'pedaço'
PESTANA	provavelmente basca
SAPO	obscura, pré-romana, talvez onomatopaica
SARÇA	pré-romano, talvez basco
SARNA	ibérica, 'areia grossa, escama'; ainda presente no basco com o sentido de 'areia'
SEARA	talvez celta
SOLAPAR	derivado de *lapa*, com o prefixo latino *sub- > sob > so*
TOUCINHO	celta, de *tucca*, 'caldo gorduroso'; só ocorre em portugalego e espanhol
TRANCA	celta, aparentada ao gaulês *tarinca*, 'espeto, trava, barrote'; só ocorre em portugalego e espanhol
VÁRZEA	obscura, pré-romana
VEIGA	talvez ibérica

Convém mencionar também as terminações *-arr(a/o)*, *-orr(a/o)*, *-urr(a/o)*, presentes em palavras como *cachorro*, *casmurro*, *caturra*, *gambiarra*, *gangorra*, *pachorra*, *pigarro* e que usamos também para formar aumentativos (*bocarra*, *cabeçorra*, *homenzarrão*) — são exclusivas da Península Ibérica e têm provável origem basca.

6.2 Herança latina: acusativo, o caso lexicogênico

O latim clássico era uma língua *sintética*, isto é, exprimia as funções sintáticas das palavras por meio de desinências, ao passo que o latim vulgar e, em seguida, as

línguas românicas são *analíticas*, isto é, exprimem as funções sintáticas das palavras mediante a ordem delas no sintagma e pelo uso de elementos como artigos e preposições:

LATIM CLÁSSICO	LATIM VULGAR	PORTUGUÊS
liber Petri	libru de Petru	o livro de Pedro

Assim é que no latim clássico literário existem tantas desinências ou flexões quantas são as funções sintáticas que uma palavra pode exercer na oração:

FUNÇÃO GRAMATICAL	PORTUGUÊS	LATIM CLÁSSICO	CASO SINTÁTICO
1. sujeito	Pedro	Petr<u>us</u>	nominativo
2. complemento restritivo	de Pedro	Petr<u>i</u>	genitivo
3. objeto indireto	a Pedro	Petr<u>o</u>	dativo
4. objeto direto	Pedro	Petr<u>um</u>	acusativo
5. vocativo	ó Pedro	Petr<u>e</u>	vocativo
6. complemento oblíquo	com Pedro	cum Petr<u>o</u>	ablativo

A tendência analítica da língua falada corrente levou a eliminar essas desinências de caso, substituindo-as pelo enrijecimento da ordem dos termos na sentença (predominantemente svc, isto é, sujeito-verbo-complemento [▶453]) e pelo uso de artigos (inexistentes em latim) e de preposições. Graças às terminações indicadoras de caso, a ordem das palavras em latim (clássico, literário) era muito mais livre: *Petrus amat Paulum, Amat Petrus Paulum, Paulum Petrus amat, Amat Paulum Petrus* etc. Todas essas frases significam "Pedro ama Paulo", mas não podemos fazer as mesmas inversões de ordem do latim, porque o sentido do enunciado muda totalmente: "Paulo ama Pedro".

O resultado dessas transformações foi que apenas dois casos restaram no latim vulgar: o nominativo e o acusativo, ou seja, um caso reto (sujeito) e um caso oblíquo (complementos). Depois dessa redução, as funções inerentes aos outros casos foram exercidas pelo acusativo com preposição. Desse modo, coube ao acusativo precedido de *de* e *ad* a expressão do genitivo e do dativo; e regida pelas preposições *de*, *per* e *cum*, a expressão do ablativo. Na Península Ibérica, o acusativo e o nominativo se fundiram, com predominância do acusativo. Daí se dizer que o acusativo, nas línguas românicas ibéricas, é o caso lexicogênico, ou seja, é da forma que as palavras tinham nesse caso sintático que se originou o léxico dessas línguas. Observe, por exemplo, que a palavra *homem* não poderia proceder do nominativo latino *homo*, mas sim do acusativo *homĭnem*. Cedo, no latim vulgar, a desinência *-m*, característica do acusativo, se perdeu, de modo que o étimo de *homem* é a forma *homĭne-*. Com a evolução da vogal breve *ĭ* em [e] e com a síncope regular do *-n-* intervocálico, surgiu *homẽẽ* que, com posterior crase dos *ẽẽ*, resultou na forma arcaica *homẽ*, que escrevemos hoje *homem*. Portanto, na reconstrução do étimo de uma palavra das línguas do grupo portugalego, é necessário recorrer à forma da palavra latina original no caso *acusativo*, de modo que para *livro* devemos recorrer a *libru(m)* e não a *liber*;

para *carne* devemos recorrer a *carne(m)* e não a *caro*; para *visão* devemos recorrer a *visione(m)* e não a *visio*.

No entanto, embora o acusativo seja o caso lexicogênico da maioria das palavras, restaram alguns vestígios, na língua, de outros casos provenientes do latim:

(a) do *nominativo*: nomes próprios como Deus, Cícero, César, Marcos, Nero, Júpiter etc. Palavras eruditas como *sóror, serpe, câncer, ladro, virgo, libido, virago* etc. Os índices pessoais do caso reto: *eu, tu, nós, vós*. Os demonstrativos *este, esse, aquele*;

(b) do *genitivo*: as palavras compostas eruditas como
terremoto (terrae + moto), 'movimento *da* terra';
aqueduto (aquae + ductu-), 'duto *de* água';
agricultura (agri + cultura), 'cultura *do* campo';
jurisprudência (juris + prudentia); 'provisão *da* justiça';
uxoricida (uxoris + cida); 'assassino *da* esposa'

(c) do *dativo*: *crucifixo* (*cruci* + *fixu* = fixado *na* cruz); os pronomes *mim (mihi), ti (tibi), si (sibi), lhe (illi)*.

No latim clássico, as palavras eram divididas, segundo a terminação, em cinco grandes classes, chamadas declinações. O que identificava uma declinação era a desinência que a palavra apresentava no caso genitivo (por isso, nos dicionários de latim, as palavras vêm sempre expressas no nominativo, seguidas da terminação do genitivo):

1ª DECL.	2ª DECL.	3ª DECL.	4ª DECL.	5ª DECL.
hora, -ae	lupus, -i	mare, -is	cantus, -us	dies, -ei

No latim vulgar, essas cinco declinações se reduziram a três, sobretudo porque eram poucos os nomes que se enquadravam na 4ª e na 5ª. Assim, os nomes da 5ª passaram, em sua maioria, para a 1ª e, em menor medida, para a 3ª declinação. Os nomes da 4ª se transferiram para a 2ª, pela semelhança entre as desinências casuais.

Para isso contribuiu a confusão que já existia no próprio latim clássico, em que alguns substantivos da 5ª podiam também ser declinados pela 1ª: *avarities, -ei* ou *avaritia, -ae*; *luxuries, -ei* ou *luxuria, -ae*; *materies, -ei* ou *materia, -ae*. O mesmo ocorria com os nomes da 4ª e da 2ª: *domus, -us* ou *domus, -i*; *colus, -us* ou *colus, -i*; *fructus, -us* ou *fructus, -i*.

A pressão analítica da mudança linguística levou, com vistas à clareza de expressão, ao emprego frequente de preposição. Onde bastava a desinência de caso para indicar a função, surgiu a partícula, facilitando a compreensão do sentido. Assim, em vez do *genitivo*, aparece, no próprio latim clássico, o *ablativo* com a preposição *de*: *nil gustabit* de *meo* (Plauto), *partem* de *istius impudentia* (Cícero). Em lugar do *dativo*, usa-se o *acusativo*, regido de *ad*: ad *me magna nuntiavit* (Plauto), ad *propinquos restituit* (Tito Lívio) etc.

Se na língua escrita mais monitorada já se verificava o fenômeno, na língua falada corrente a necessidade de clareza, maior e mais urgente, tornou categórico o emprego da preposição. Com isso, a maior parte dos casos, tornados desnecessários, desapareceu. Restaram apenas, como já se mencionou, o *nominativo* (caso do sujeito) e o *acusativo* (caso do objeto). Nas línguas românicas ibéricas, houve a fusão desses dois casos. No francês arcaico, permaneceu durante algum tempo a distinção entre dois casos, o caso-sujeito e o caso-regime (para todas as funções diferentes do sujeito), que não sobreviveu. Na língua romena, até os dias de hoje, se conserva a distinção entre dois casos (um nominativo/acusativo, e outro, dativo/genitivo).

6.3 O GÊNERO DOS SUBSTANTIVOS: DESAPARECIMENTO DO NEUTRO

Pode ser que no indo-europeu primitivo o gênero gramatical dos nomes se fundamentasse no sexo biológico real. Por isso, os seres inanimados eram do gênero neutro (*ne uter*, "nem um nem outro"). Essa distinção, no entanto, logo perdeu todo vínculo com a realidade objetiva, e o gênero se tornou uma categoria exclusivamente gramatical e, portanto, arbitrária. Por causa disso, encontramos no grego e no latim, por exemplo, seres inanimados classificados como masculinos ou femininos. Os substantivos abstratos em latim podiam ser de qualquer gênero: *religio, religionis* é feminino, enquanto *amor, amoris* é masculino, e *examen, examinis* é neutro.

Na transformação do latim nas línguas românicas, o gênero neutro aos poucos desapareceu e deixou de existir como categoria gramatical. Vamos ver os fatores que provocaram o desaparecimento do gênero neutro:

1. Na 1ª declinação não existiam nomes neutros: eram quase todos femininos, de tal forma que a terminação *-a* passou a ser característica dos nomes femininos nas línguas românicas ibéricas, itálicas e no romeno, ao contrário do latim, em que as palavras femininas podiam ter as mais diversas terminações (inclusive *-o*, como no nominativo de *passio, religio, virgo, vertigo* etc.).

2. Na 2ª declinação, a maioria dos nomes eram masculinos e neutros. Com isso, a terminação *-o* (do acusativo singular *-um* > *-u* > *-o*) se tornou a característica dos nomes masculinos nos romances ibéricos, itálicos e no romeno. Os substantivos neutros, como tinham suas desinências idênticas às dos masculinos, também passaram a esse gênero: *pratum* > *pratu* > *prado*; *exemplum* > *exemplu* > *exemplo*; *templum* > *templu* > *templo*; *vinum* > *vinu* > *vĩo* > *vinho* etc.

No entanto, como a terminação do plural dos neutros era *-a* (*exemplum — exempla*), ocorreram confusões desse plural com o gênero feminino. É o que se verifica com as palavras usadas com o valor de pluralidade ou de coleção que, neutras plurais em latim, se transformaram em femininas singulares em português:

LATIM		PORTUGUÊS
NEUTRO SINGULAR	NEUTRO PLURAL	FEMININO SINGULAR
ovum	ova	ova
folium	folia	folha
brachium	brachia	braça
lignum	ligna	lenha
interaneum	interanea	entranha
vestimentum	vestimenta	vestimenta

3. Os nomes neutros da 3ª declinação passaram, em geral, para a 2ª, assumindo o gênero masculino. Alguns, cujo gênero já alternava em latim com o masculino ou o feminino, passaram ao português com esses gêneros: *mare* > *mar* (fem. no port. arc. e masc. no moderno); *rete* > *rede* (fem.).

4. O gênero neutro só existe hoje no romeno. Nas demais línguas românicas, ele desapareceu. No entanto, sobreviveram alguns vestígios dele na língua atual:

   (a) demonstrativos — *isto, isso, aquilo*
   (b) quantificadores indefinidos — *tudo, nada, algo*;
   (c) adjetivo substantivado — *o útil, o agradável, o belo*;
   (d) infinitivo substantivado — "dirigiu *o olhar* para mim"; "*fumar* é prejudicial à saúde"; "sede de *poder*"

6.4 Adjetivos

A morfologia flexional dos adjetivos no latim vulgar era menos complexa do que a dos substantivos, já que era regulada pelas desinências das declinações que se mantiveram na língua falada corrente.

Os adjetivos chamados de 1ª classe na gramática latina tinham uma forma destinada a se adaptar a cada um dos gêneros dos substantivos com os quais concordavam: *paucus* (masc.), *pauca* (fem.), *paucum* (neutro). Depois que o neutro desapareceu e os casos foram eliminados, os adjetivos assumiram as seguintes formas:

SINGULAR: -u > -o, -a > -a: *paucu* > *pouco*, *pauca* > *pouca*;
PLURAL: -os > -os, -as> -as: *paucos* > *poucos*, *paucas* > *poucas*.

Com o desaparecimento da forma do neutro, acrescido ao fato de serem idênticas as desinências do masculino e do feminino no único caso sobrevivente (o acusativo), os adjetivos de 2ª classe, que antes podiam ser triformes, biformes ou uniformes, acabaram por ficar todos uniformes, qualidade que persiste nos seus correspondentes em português:

SINGULAR: -e > -e: *celebre* > *célebre*; *leve* > *leve*; *cortense* > *cortês*
PLURAL: -es > -es: *celebres* > *célebres*; *leves* > *leves*; *cortenses* > *corteses*

A erosão do sistema de casos e a transformação das terminações *-a* e *-o* em verdadeiros morfemas de gênero fez com que as palavras que passaram a terminar em *-e* não trouxessem marca de gênero explícita. Para preencher essa lacuna, os falantes passaram a usar o demonstrativo ou o numeral: *illa fonte*, *unu dente*.

Daí surgiu, nas línguas românicas, o artigo, inexistente em latim clássico: *a fonte*. (Como veremos em [▶782], não existe de fato o *artigo indefinido*). Quando, mais tarde, diversas palavras, por terem sofrido apócope [▶296], passaram a terminar em consoante (*anel*, *cruz*, *luz*, *mar*, *mês*, *voz* etc.), o uso do artigo se tornou ainda mais necessário.

6.5 Relatinização, convergência e divergência

No período renascentista, quando foram produzidas as primeiras gramáticas do português e se formou o vocabulário **erudito** da língua — termos cunhados para uso literário, científico, teológico, filosófico, técnico etc. —, ocorreu um fenômeno que hoje designamos como **relatinização**. Esse nome se explica porque, na criação daquele vocabulário, os letrados recorreram ao latim clássico. Esse processo não afetou somente o vocabulário, mas igualmente a gramática, na medida em que se tentava reintroduzir no português erudito construções sintáticas próprias do latim literário.

Essas novas palavras eruditas constituem verdadeiros *estrangeirismos*, uma vez que não sofreram os processos de mudança linguística que afetaram o léxico geral da língua. Nelas tudo o que se fez foi um *aportuguesamento* mínimo, de tal modo que entre o termo erudito e o étimo latino a diferença fosse quase imperceptível, como podemos ver no quadro a seguir:

LATIM	PORTUGUÊS			
	FORMA ERUDITA	DATAÇÃO	FORMA POPULAR	DATAÇÃO
dígitu-	dígito	1532	dedo	séc. XIII
cáthedra-	cátedra	séc. XIV	cadeira	952
duplu-	duplo	1651	dobro	1277
frígidu-	frígido	1542	frio	1101
íntegru-	íntegro	séc. XIV	inteiro	1093
legítimu-	legítimo	séc. XV	lídimo	1328
límpidu-	límpido	1572	limpo	séc. XIII
mácula-	mácula	1589	malha	séc. XIII
mágicu-	mágico	séc. XIV	meigo	1175
másculu-	másculo	1687	macho	séc. XIII
matéria-	matéria	séc. XIV	madeira	1269
óculu-	óculo	1649	olho	séc. XIII
plaga-	plaga	séc. XIV	chaga	1297
planu-	plano	séc. XIV	chão	1261
plenu-	pleno	1708	cheio	séc. XIII
primáriu-	primário	1685	primeiro	1093
rígidu-	rígido	1572	rijo	séc. XIII
sede-	sede	1446	sé	séc. XIII

Além da formação de vocábulos novos, ocorreu também uma tentativa de *recondução* de palavras, já muito antigas na língua, a uma forma mais próxima do étimo latino. Algumas dessas reconduções tiveram êxito, como as seguintes:

FORMA TRADICIONAL	FORMA RELATINIZADA
avondança	abundância
chantar	plantar
coa	cauda
esmar	estimar
fremoso	formoso
frol	flor
goivo	gáudio
letradura	literatura
marteiro	martírio
obridar	olvidar
pendença	penitência
seenço	silêncio

Todo esse processo de *intervenção consciente* no léxico da língua fez aumentar a quantidade das chamadas formas divergentes — palavras que remontam a um único étimo. Existem casos de formas divergentes de formação popular (não erudita), em que de um mesmo étimo os falantes derivaram mais de um vocábulo:

régula-	régua regra relha	*capitale-*	capital cabedal caudal

Com o processo de relatinização, como vemos no quadro acima, muitas formas divergentes novas surgiram, sempre na oposição entre formação popular e formação erudita:

clamare	chamar clamar	*clave-*	chave clave	*coágulu*	coalho coágulo	*cómputu*	cômputo conto

Uma das evidências mais claras do caráter "não popular" dessas formas divergentes novas é que muitas delas têm acentuação proparoxítona (*dígito, másculo, límpido, rígido, coágulo, cômputo* etc.), contrariando a tendência mais recorrente da língua, que é a acentuação paroxítona (*dedo, macho, limpo, rijo, coalho, conto* etc.).

Também existem, ao contrário, formas convergentes: uma mesma forma em português que remete a mais de um étimo latino, como nos exemplos abaixo:

cooperet ('ele cobre') *cupre-* (metal)	COBRE	*filare* ('tecer fio') *fidare* ('ter fé')	FIAR	*cito* (adv.) *cedo* (v.)	CEDO

No entanto, como fenômeno de interesse para o estudo da mudança linguística, as formas divergentes têm importância muito mais destacada.

6.6 Empréstimos e estrangeirismos

Como toda e qualquer língua viva do mundo, a nossa também recebeu, tem recebido, recebe e receberá contribuições lexicais de diferentes línguas. Essas contribuições são tradicionalmente classificadas como empréstimos, um termo não muito adequado, já que *emprestar* implica na ideia de *devolver*, coisa que não ocorre nesse caso. Vimos acima que diferentes línguas nos legaram palavras que já deviam circular pela Península Ibérica antes da chegada dos romanos e da imposição do latim. Depois da romanização da Península, outros povos vieram se instalar ali como os de línguas germânicas e árabe [▸214], que também deixaram suas marcas no nosso patrimônio lexical. Fenômenos históricos, culturais, políticos, ideológicos etc. se com-

binam para que palavras de outras línguas e, principalmente, da língua dominante no plano internacional, se instalem no léxico de outras. Sempre foi assim e com a intensificação das relações internacionais e das tecnologias de comunicação, isso se torna ainda mais frequente e rápido nos dias de hoje.

Recentemente no Brasil (mais precisamente a partir de 1999, quando um deputado federal tentou fazer aprovar uma lei contra o uso de termos oriundos do inglês), o debate em torno dos estrangeirismos voltou à cena. Como tudo o que diz respeito à língua, o tema acendeu paixões, principalmente as de cunho nacionalista, e com elas muitos argumentos irracionais. É claro que todas as pessoas têm o direito de se manifestar a respeito da língua, que é um elemento primordial da identidade individual e coletiva. Do ponto de vista político, a recusa do estrangeirismo é perfeitamente justificável.

Mas, como todo fenômeno linguístico, é preciso analisar a questão dos estrangeirismos com um mínimo de imparcialidade, a fim de se formar uma ideia mais equilibrada acerca do fato. O uso de palavras estrangeiras — mais especificamente, nos dias atuais, de palavras vindas do *inglês* — não afeta a gramática da língua, uma vez que os empréstimos são contribuições ao léxico da língua e não ao modo como a língua se estrutura morfossintaticamente nem fonologicamente. Ao contrário, quando um termo estrangeiro passa a ser empregado, ele é imediatamente adaptado aos hábitos fonéticos da língua de chegada para, mais adiante, se permanecer em uso, receber uma grafia coerente com o modo convencional de escrever o idioma.

Foi assim que o francês *abat-jour* se transformou no *abajur*, e que o *billet*, o *bonnet*, o *maillot* e o *haut-bois* se transformaram em *bilhete*, *boné*, *maiô* e *oboé*. E é assim que muitos termos provenientes do inglês estão aos poucos se fonologizando-lexicalizando-morfossintaticizando no PB. Os verbos formados com radicais da língua inglesa são sempre da 1ª conjugação, que é aquela que, em todas as línguas românicas, serve de paradigma para a criação de novos verbos. Por isso é que temos *deletar*, *printar*, *shrinkar*, *palletizar* e outros verbos, usados principalmente em áreas técnicas específicas. E é por isso que pronunciamos a palavra *e-mail* como [i'meɪu] com três sílabas, quando em inglês sua pronúncia é [i'mɛɬ], com duas sílabas e realizações fonéticas bem diferentes.

Qualquer enunciado feito em PB com palavras de origem estrangeira vai se construir de acordo com as regras morfossintáticas da língua:

> Eu trouxe o meu *notebook*, mas esqueci o teu *pen-drive* na mesa do *coffee-break*...

A ordem dos constituintes no enunciado, a concordância de gênero e número, as preposições, a flexão verbal... tudo é o mais puro e legítimo PB. Somente os nomes revelam sua origem estrangeira. Um falante nativo de inglês jamais compreenderia esse enunciado.

É preciso, então, distinguir os dois aspectos da problemática do estrangeirismo. O aspecto *sociocultural*, que varia de pessoa para pessoa e de grupo para grupo — com

alguns segmentos da sociedade valorizando o emprego de termos estrangeiros como sinal de "progresso" e "sofisticação", enquanto outros o rejeitam em nome de uma preservação de valores culturais e identitários nacionais — e o aspecto propriamente *linguístico*, em que o uso das palavras não afeta nem compromete a sobrevivência da língua receptora. Por isso, não é possível aceitar, como escreveu o mencionado deputado em seu projeto de lei, o argumento de que a população brasileira está para se tornar "bilíngue". É esquecer o grande contingente de analfabetos plenos e analfabetos funcionais da nossa população, milhões de pessoas que desconhecem completamente a língua inglesa, e desconhecer os reais mecanismos socioculturais e sociolinguísticos que atuam na circulação e no emprego de termos estrangeiros.

6.7 Herança latina: os prevérbios

Chamamos de **prevérbios** os prefixos que se unem a uma raiz para formar novas palavras. Em latim, a maioria dos prevérbios eram preposições que, ao se juntar às raízes, conferiam a elas novos significados e sentidos. Muitos desses prevérbios apresentam uma consoante final que, como vamos ver, acaba se **assimilando** à consoante inicial da raiz, formando uma consoante dupla. No quadro abaixo, vamos listar esses prevérbios com seu significado principal. Observe que a maioria deles permanece em uso no portugalego, já que é deles que provieram muitas das nossas preposições. Perceba também de que modo o significado original, o primeiro da lista, vai se desdobrando em outros, derivados dele, pelo processo de **metaforização** que estudamos em [▶170], como a passagem do sentido de *lugar/espaço* para o de *tempo/processo*.

PREVÉRBIO	SIGNIFICADO
ab- (a-; abs-; au-)	'afastamento'; 'separação'; 'recusa'; 'negação'; 'anulação' etc.
ad- (a-)	'aproximação'; 'acréscimo'; 'intensidade'; 'início da ação' etc.
ambi- (amb-)	'em torno', 'em volta', 'dos dois lados', 'de vários lados'
ante-	'posição frontal'; 'na frente'; 'antes de'
bi-	'dois'
circum- (circu-)	'movimento circular'
cis-	'aquém de'; 'da parte de cá'
com- (co-)	'convergência'; 'reunião'; 'companhia'; 'ao mesmo tempo' etc.
contra- (contro-)	'frente a frente'; 'oposição'; 'hostilidade'
de-	'queda', 'descida', 'de cima para baixo'; 'afastamento'; 'diminuição' etc.
dis- (di-)	'dispersão'; 'afastamento'; 'divergência'; 'distribuição' etc.

ex- (e-)	'saída'; 'de dentro para fora'; 'para cima'; 'privação' etc.
extra- (extro-)	'que está fora de'
in-[1]	'negação'
in-[2]	'entrada'; 'para dentro'; 'aproximação'; 'hostilidade'; 'mudança de estado'
inter-	'entre'; 'de vez em quando'; 'separação de algo que era contínuo' etc.
intra- (intro-)	'entrada'; 'para dentro'
ob- (o-; obs-)	'na frente de'; 'oposição';
paene-	'quase'
per-	'através de'; 'duração'; 'intensidade'; 'completude'; 'degradação' etc.
post-	'atrás' no espaço e 'depois' no tempo
prae-	'na frente', 'antes'; 'superioridade'; 'intensidade' etc.
praeter-	'para o outro lado'; 'passado'
prim-	'na frente', 'antes', 'primeiro'
pro-	'para a frente', 'na frente de todos'; 'descendência'; 'substituição'
re- (red-)	'para trás', 'retorno'; 'repetição'; 'recusa'; 'anulação' etc.
retro-	'para trás'
se- (sed-)	'afastamento', 'privação'
semi- (se-; ses-)	'metade'
sesqui-	'um e meio'
sub- (sus-)	'de baixo para cima', 'embaixo', 'para baixo'; 'submissão'; 'substituição'
subter-	'por baixo'
super-	'por cima'; 'abundância'; 'superação' etc.
supra-	'acima'
trans- (tra-; tran-)	'para o outro lado'; 'mudança de estado'
tri-	'três'
ultra-	'para além de'; 'da parte de lá'
uni- (un-)	'um'

Com esses prevérbios era possível, em latim, formar inúmeros derivados, enriquecendo o léxico do idioma. Nas línguas românicas, continuamos a fazer isso, embora muitas vezes tenhamos herdado a palavra diretamente do latim, de modo que raramente percebemos o *étimo* principal, isto é, a raiz da palavra que, por ter sofrido processos de mudança fonética, se tornou irreconhecível.

Quando estudamos inglês, ficamos admirados com a possibilidade de usar as preposições para obter *sentidos* diferentes quando vêm *pospostas* aos verbos. Por exemplo, o verbo *look*, que tem o sentido mais geral de 'olhar', ganha novos sentidos conforme a preposição que se coloca depois dele:

- *look after* → 'cuidar, tomar conta de'
- *look for* → 'procurar, buscar'
- *look forward* → 'ansiar'
- *look upon* → 'considerar, estimar' etc.

Ora, a diferença entre o latim e o inglês, nesse caso, é que, no latim, as partículas eram colocadas *antes* do verbo (por isso, *prevérbios*), enquanto em inglês elas se colocam *depois*:

latim: *exire* ('fora' + 'ir' = 'sair')
inglês: *go out* ('ir' + 'fora' = 'sair')

No latim, contudo, essa colocação do prevérbio antes da raiz levava a algumas transformações fonéticas importantes. A primeira e mais comum era a assimilação, [▶299]: quando um prevérbio terminava em consoante e a raiz começava em consoante, a consoante do prevérbio era assimilada à da raiz, gerando uma consoante dupla (que era pronunciada em latim clássico como dupla):

ad + *currĕre* = *accurrĕre* ('acorrer')
ob + *mittĕre* = *ommittĕre* ('omitir')
sub + *ponĕre* = *supponĕre* ('supor')

Outro fenômeno fonético comum nessas composições era a apofonia. Você já se perguntou por que o contrário de *fácil é difícil*, e não "disfácil"? Frequentemente, quando a raiz continha uma vogal aberta, a afixação de um prevérbio causava o fechamento dessa vogal em [e] ou [i], conforme os casos. Por isso, para o observador leigo, é difícil entender que palavras como *agir* e *exigir* tenham a mesma raiz *ag-*: o prevérbio *ex-* levou ao fechamento do [a] em [i]. Observe mais alguns exemplos:

amicus ('amigo')	*in- -amicus = inimicus* ('inimigo')
aptus ('apto')	*in- -aptus = ineptus* ('inepto')
barba ('barba')	*in- -barba = imberbe* ('sem barba, imberbe')
cadens ('cadente, que cai')	*ob- -cadens = occidens* ('ocidente')
candens ('branco fulgurante')	*in- -cand- = incendium* ('incêndio')
cantus ('canto')	*ad- -cantus = accentus* ('acento')
facilis ('fácil')	*dis- -facilis = difficilis* ('difícil')
gradus ('passo, degrau')	*pro- -grad- = progrido* ('progrido')
habeo ('tenho')	*pro- -habeo = prohibeo* ('proíbo')
tango ('eu toco')	*ad- -tango = attingo* ('eu atinjo')

Para termos uma pequena ideia do que era possível fazer em termos de derivação com esses prevérbios, vamos examinar aqui as palavras compostas com base na raiz *ten-*, origem do verbo *tenēre* ('ter'). Não vamos listar as que existiam em latim, muito mais numerosas, mas somente aquelas que permanecem vivas em português:

TEN-
abstenção, abstencionismo, abster, abstinência, abstinente...
ater, atinência, atinente...
contenedor, contenença, contentador, contentamento, contentar, contente...
conter, conteudismo, conteudista, conteudístico, conteúdo...
continência, continental, continente...
continuação, continuar, continuidade, continuísmo, continuísta, continuístico, contínuo...
detenção, detentor, detençoso, detenedor, detenimento, detenível, detento, deter...
entretenimento, entreter, entretido, entretimento...
extenuação, extenuado, extenuador, extenuante, extenuar, extenuativo, extenuável...
impertinência, impertinenciar, impertinente...
manutenção, manter, mantenedor, mantimento... (de *manu-*, 'mão', + 'ter')
obtenção, obtenível, obtentor, obter...
pertença, pertence, pertencença, pertencente, pertencer, pertencido, pertencimento...
pertinaz, pertinácia, pertinacidade, pertinacíssimo, pertinência, pertinente
portento; portentosidade, portentoso
retenção, retência, retencionário, retencionista, retenir, retentiva, reter, retido, retinência...
sustentar, sustenido, sustenir, sustenizar, sustentável, sustentação, sustentar, suster...

Sem prevérbios, temos muitos derivados dessa raiz também: *tenaz*, *tenacidade*, *tenente*, *tenência*, *tênis*, *tenista* etc.

Alguns verbos latinos não sobreviveram no português moderno, mas seus derivados permanecem muito ativos na língua. É o caso de *ducĕre*, 'levar, conduzir', que deu, no português antigo, *duzir*, verbo que logo desapareceu do uso normal, e *pellĕre* ('empurrar, puxar'), que nos deu *-pelir*; seu particípio passado era *pulsu-*, que também gerou muitas palavras na nossa língua. De um derivado seu, *pulsare*, surgiu o nosso *puxar*, também muito produtivo. O mesmo vale para o verbo *ferre* ('levar, trazer') que se transformou em *ferĕre*, gerando o português *-ferir* (sem nenhum parentesco com o verbo *ferir*, derivado de outro verbo latino). Seus derivados, porém, são muitos e muito empregados:

UNIVERSIDADE FEDERAL Biblioteca Comunitária

-DUZIR	abduzir • aduzir • conduzir deduzir • induzir • introduzir produzir • reduzir • reproduzir seduzir • traduzir
-FERIR	aferir • auferir • conferir deferir • desferir • diferir indeferir • inferir • oferecer preferir • proferir • referir sofrer • transferir
-PELIR	compelir (compulsão etc.) expelir (expulso etc.) impelir (impulso etc.) propelir (propulsão etc.) repelir (repulsa etc.)

Outro verbo latino que não se preservou foi *struĕre* ('levantar, erguer, amontoar'). Mesmo assim, ele nos legou uma família muito importante: *construir, destruir, instruir, obstruir, reconstruir* e, também, *indestrutível, indústria, instrumento*.

O verbo latino *stāre* ('estar de pé') se tornou um dos mais usados da nossa língua, principalmente por fazer par com *sedēre* ('estar sentado'), criando a dupla ser/estar, tão importante para nós, que fazemos questão da diferença entre *ser feliz* e *estar feliz*, enquanto muitas outras línguas têm um verbo só (inglês, francês, alemão etc.) e outras, verbo nenhum (russo, tcheco, quéchua, línguas semíticas, húngaro, chinês etc.). Do sentido inicial de 'estar de pé' deriva a noção de algo que é 'temporário, passageiro, fugaz' etc.

O aprendizado dos usos de *ser* e *estar* por estrangeiros não é coisa fácil, porque algumas vezes a lógica se inverte: dizemos, por exemplo, que alguém está morto quando esse estado não tem nada de fugaz e passageiro... Nem mesmo os falantes de espanhol, língua em que também existe a dupla *ser/estar* (e, por conseguinte, nós quando vamos falar espanhol), deixam de cometer equívocos ao usar os mesmos verbos em PB: afinal, em espanhol se diz, por exemplo, que uma mesa *está hecha de madera*, quando nós dizemos que ela é *feita de madeira*. O título da canção, brasileira, "*Soy loco por ti América*" não faz sentido em espanhol, onde o que se diria é."Estoy loco por ti, América..."

Esses verbos são chamados *cópulas* ou *verbos de ligação* e seu significado é tão básico e fundamental que eles podem ser dispensados, como em tantas línguas, por estarem desde sempre subentendidos quando duas palavras são postas lado a lado para formar a relação sujeito-predicado. Como veremos em [▸407], Platão considerava os adjetivos como verbos porque neles o verbo *ser* está sempre subentendido (*homem gentil = homem que é gentil*).

Com a raiz do verbo *stāre* se formaram muitos verbos (para não mencionar as incontáveis palavras de outras classes). Essa mesma raiz, no entanto, já em latim, passou

por modificações que fizeram surgir novas raízes, todas também muito produtivas nos processos de derivação: com redobro do presente temos *sisto* (infinitivo: *sistĕre*; part. pas.: *stătu-*), a partir do particípio passado (*stātu-*) temos *statuo* (infinitivo: *statuēre*; part. pas.: *statutu-*) e com alongamento da vogal a raiz *stau-*, só usada em compostos. Vamos organizar isso tudo num quadro para ver melhor:

ST-	arrestar circunstar constar constatar contrastar custar distar estabelecer estar instar obstar prestar restar **[Obs.: os radicais *subst-* e *superst-* não nos deram verbos, mas uma série de palavras muito usadas (*substância*, *superstição* etc.)]**
STAT-	constituir destituir instituir prostituir restituir substituir **[Obs.: em todos esses derivados ocorreu apofonia, com fechamento do *a* em *i*. Desse verbo provém a palavra *estátua*, entre tantas outras.]**
SIST-	assistir consistir desistir destinar exsistir insistir obstinar persistir resistir subsistir **[Obs.: os verbos *destinar* e *obstinar* são formações mais tardias, por isso a diferença de radical.]**
STAU-	instaurar restaurar

E já que falamos de *stare*, é mais do que justo examinar também seu parceiro *sedēre*, fonte do nosso verbo ser. O significado do verbo em latim era 'estar sentado' e daí, por metáfora, o sentido de algo que permanece, que dura, que existe, que... *é*. O particípio passado de *sedēre* era *sessu-*, o que explica a existência de diversas palavras em PB com esse radical também, além de *sed-*. Em todas essas

palavras está presente a ideia de 'fixação, estabelecimento, assentamento' etc. Vamos conhecer algumas delas:

SED-	assédio - assediar assentar (ver sentar, abaixo) obsedar - obsedante sé (< sede) sede - sediar sedentário sedimento sedar - sedativo sela (< sella < sedla) sentar (< seentar < *sedentare)
SID-	assíduo desídia dissidente insídia - insidioso presidente - presidir presídio residir - residência resíduo subsídio **[Obs.: radical com apofonia *e* > *i*].**
SESS-	assessor - assessorar assessório circunsessão desassossego (v. sossego, abaixo) obsessão sessão sessionar sossego (< sessego)

Conforme veremos [▶609], o verbo correspondente ao nosso *ser* era, em latim, *esse*, um verbo extremamente irregular desde suas remotas origens no indo-europeu. Ele se manteve em italiano (*éssere*) e em francês (*être*), mas nas línguas ibéricas foi substituído pelo *sedere*, que acabamos de ver. Dele, porém, derivam algumas palavras muito empregadas até hoje: *essência, essencial, interesse* e sua família. O particípio presente era *ente-*, que nos deu *ente, entidade, ausente* (abs.*entem*), *presente* (praes. *entem*) e toda a sua família (*apresentar, presenciar, representar* etc.)

Por fim, vamos mencionar outros dois verbos que apresentam uma grande quantidade de derivados: o verbo *pôr* e o verbo *meter*. Esse segundo verbo, no PB atual, teve seu significado restringido à noção de 'pôr, introduzir', mas seu étimo latino, *mittĕre*, significava 'deixar ir, deixar partir, soltar, largar, lançar, atirar, enviar'. Observe que os verbos derivados oscilam entre as formas *-meter* e *-mitir* (nesse caso, não se trata de apofonia, mas da transformação regular do [ĭ] latino no nosso [e] fechado [▶302]):

-METER /-MITIR	-POR
acometer	apor
admitir	compor
arremeter	contrapor
cometer	decompor
comprometer	depor
demitir	dispor
emitir	expor
imitir	impor
intermitir	indispor
intrometer	interpor
manumitir	justapor
omitir	opor
permitir	predispor
prometer	propor
remeter	recompor
submeter	repor
transmitir	sobrepor
	superpor
	supor
	transpor

Convém lembrar que o particípio passado do verbo latino *mittĕre* era *missu-*, de onde provém uma grande quantidade de palavras nossas (ora com *miss-*, ora com *mess-*): *arremesso, arremessar, demissão, missa, missionário, missão, promessa, comissão, remessa* etc.

Por seu lado, o particípio passado do verbo *ponĕre* (> *poer* > *pôr*) era *posĭtu-*, que se transformou em *posto*. Os radicais *post-* e *pósit-* (formação erudita) estão presentes em *aposto, apostar, depósito, depositar, posição, compósito, compostar, postar, repositório, supositório* etc.

6.8 Infixo nasal

Em latim também ocorria um fenômeno fonético interessante que explica certas divergências nas formas derivadas de uma mesma raiz: é o chamado infixo nasal. Em diversos verbos, as formas conjugadas do presente apresentavam um -N- (ou -M-) que se inseria dentro da raiz (daí o nome in-fixo). Nas demais formas verbais, porém, esse infixo não aparecia:

TAG- : raiz
taNgo ('eu toco') - presente do indicativo
tetigi ('eu toquei') - pretérito perfeito
tactum ('tocado') - particípio passado
taNgĕre ('tocar') - infinitivo

Com isso, os derivados oscilam, ora apresentando o infixo, ora não apresentando. Vejamos alguns exemplos:

cub- ('estar deitado')	*cub-* *cumb-*	concubino, cúbito, decúbito, incubar, íncubo incumbir, sucumbir
fig- ('modelar, esculpir')	*fig-* *fingo* *fictum*	figura, figurar, figurante fingir, fingimento, finta ficção, fictício
frag- ('quebrar')	*frag-* *fractum* *frango*	frágil, fragmento, fragorar, fragoroso, fragoso fra(c)ção, fra(c)cionar, fra(c)tura, fra(c)turar frangalho, confranger
pig- ('pintar')	*pig-* *pingo* *pictum*	pigmento, pi(g)menta pintar, pintor, pintura pictórico, pi(c)toresco
rup- ('romper')	*rumpo* *ruptum*	rompante, romper, rompimento ruptura
sac- ('sagrar')	*sacrum* *sancio* *sanctum*	sacro, sacrário, sacrilégio, sagrar, sagrado sancionar santo, santuário, santidade
tag- ('tocar')	*tango* *tactum*	tangência, tangencial, tangente, tanger ta(c)tear, tá(c)tico, tá(c)til, ta(c)to
vic- ('vencer')	*vinco* *victum*	vencedor, vencer, vencimento, vencível ví(c)tima, vi(c)tória, ví(c)tor

O verbo *tango* ilustra bem os três fenômenos fonéticos que acabamos de resenhar:

- *ad- -tangere* > *attingere* ('atingir') → assimilação, apofonia e infixo
- *con- tag-* > *contiguus* ('contíguo') → apofonia
- *con- -tang-* > *contingens* ('contingente') → infixo e apofonia
- *in- -tag-* > *integru-* ('intacto, inteiro') → apofonia

6.9 Raízes enterradas no tempo

6.9.1 Falamos indo-europeu?

As palavras que usamos diariamente remontam a tempos muito distantes — na verdade, remontam à pré-história. Não estamos falando simplesmente do Império romano e do latim, mas da língua da qual o próprio latim se originou: o protoindo-europeu. Como essa língua deve ter sido falada há mais de 5.000 anos por um povo sem cultura escrita, o que sabemos dela é o resultado do trabalho de *reconstrução* que vem sendo feito há duzentos anos pelos linguistas dedicados a ele. Essa reconstrução se faz por meio da *comparação* entre as diferentes línguas da família e, com base nessa comparação, os estudiosos propõem *raízes* hipotéticas (marcadas com um asterisco [*]) das quais teriam se originado as palavras existentes nas línguas conhecidas.

Os falantes mais remotos do protoindo-europeu devem ter habitado uma região situada na Ucrânia atual. Essa localização, contudo, está sempre sujeita a debates, de modo que permanece como uma hipótese. A partir desse centro original, provavelmente impelidas pelo aumento da população e pela escassez de alimento, ondas migratórias partiram em diferentes direções, tanto para o leste, rumo à Índia, quanto para o oeste, rumo à Europa ocidental. As migrações rumo ao oriente explicam a existência de línguas indo-europeias nas montanhas do Cáucaso (armênio), na Ásia Menor (atual Turquia: o hitita, falado num império que guerreou contra o Egito dos faraós), na Pérsia (atuais Irã e Afeganistão), na Índia e no sudoeste da China.

Como disse Ferdinand de Saussure, o que nós falamos hoje é, numa elevada porcentagem, pura herança indo-europeia. O nome indo-europeu foi dado a esse grupo pelos estudiosos do século XIX porque as línguas que o compõem são faladas desde o extremo ocidental da Europa (Portugal e Irlanda) até mais da metade da Índia. Com a expansão marítima e colonial das nações europeias, algumas línguas indo-europeias foram levadas para outros continentes (português, espanhol, francês, inglês, holandês), aumentando ainda mais o número de seus falantes. No ano de 1908, foram descobertos documentos escritos em duas línguas até então desconhecidas, que receberam o nome de tocariano A e tocariano B e foram identificadas pelos estudiosos como pertencentes ao grupo indo-europeu — eram faladas em áreas que correspondem hoje à província chinesa de Xinjiang. Os documentos foram preservados intactos graças ao clima seco da região. Estão escritos em variantes do alfabeto brâmane e do alfabeto maniqueu. A descoberta revela, portanto, que os falantes do protoindo-europeu foram até além da Índia em suas migrações.

De forma esquemática, podemos representar do seguinte modo as línguas indo-europeias:

PROTOINDO-EUROPEU

GRUPO	LÍNGUAS
ALBANÊS	albanês arcaico* → guegue, tosque
ANATÓLICO	hitita*
ARMÊNIO	armênio arcaico* → armênio moderno
BÁLTICO	prussiano* → letão, lituano
ESLÁVICO	eslavônio* → búlgaro, russo, ucraniano, tcheco, polonês, esloveno...
GERMÂNICO	gótico* → alemão, sueco, inglês, holandês, islandês, dinamarquês...
HELÊNICO	grego antigo* → grego moderno
INDO-ARIANO	sânscrito* → híndi-urdu, romani, nepalês, gujarati, marati...
IRANIANO	persa antigo* → farsi, pashto, baluque, mazandarani...
ITÁLICO	osco* — umbro* — falisco* — latim* → italiano, francês, occitano, sardo, catalão, romanche, romeno, espanhol, galego, português...
TOCARIANO	tocariano A*, tocariano B*

[Obs.: o asterisco indica língua extinta.]

Essa tabela exibe os grupos e línguas bem conhecidos e documentados. Além deles, muitas outras línguas classificadas como indo-europeias desapareceram, deixando pouquíssimos indícios capazes de permitir um conhecimento mais aprofundado: ilírico, venético, libúrnio, messápio, frígio, paiônio, trácio, dácio, lígure, lusitano, macedônio antigo.

É sempre surpreendente comparar as línguas indo-europeias e ver o quanto de comum elas têm em seu vocabulário ancestral como os números, os graus de parentesco, os astros, alguns animais e plantas etc. Vamos ver algumas dessas afinidades em línguas que deixaram de existir há muito tempo. Escolhi o sânscrito, por ser a mais oriental das línguas indo-europeias com grande cultura escrita, e o irlandês antigo, por ser a mais ocidental de todas, ambas separadas entre si por dezenas de milhares de quilômetros que, na pré-história, eram uma distância que jamais permitiria qualquer contato entre os povos falantes dessas línguas. Para efeito de comparação, incluí o latim, fonte mais remota da nossa língua:

SÂNSCR.	IRL. ANT.	LATIM
bhratar-	**brathair**	**frater** ('irmão')
dva(u)	**da**	**duo** ('dois')
matar-	**mathir**	**mater** ('mãe')
pitar-	**athir**	**pater** ('pai')
sal-ilá	**salann**	**sal** ('sal')
saptá	**sech**	**septem** ('sete')
sás	**sé**	**sex** ('seis')
svaser	**siur**	**soror** ('irmã')
trayah	**tri**	**tres** ('três')
t(u)vam	**tu**	**tu**

SÂNSCR.	IRL. ANT.	LATIM
ásti	**is**	**est** ('ele é)
dan, dantam	**det**	**dens, dentis** ('dente')
daśa	**deich**	**decem** ('dez')
laghui-	**laigiu**	**levis** ('leve')
madhya	**mid**	**medius** ('meio')
mahant-	**mochtae**	**magnus** ('grande')
nama	**ainmm**	**nomen** ('nome')
nava	**noi**	**novem** ('nove')
rudhirá	**rúad**	**ruber** ('vermelho')
sátsi	**saidid**	**sedet** ('está sentado')

Acompanhar os desdobramentos de forma e de sentido que uma mesma raiz indo-europeia sofreu ao longo dos milênios nos revela inúmeras surpresas:

PROTOINDO-EUROPEU: *WID- (*WEID-, *WOID-) 'VER, SABER'	
sânscrito:	**vid** ('descobrir, compreender, saber') **vidya** ('conhecimento, aprendizagem, ciência') **avidya** ('ignorância, desconhecimento') **veda** ('conhecimento ritual, saber secreto')
grego:	**(w)idéa** ('aparência, tipo, natureza, forma etc.') **eidon** ('eu vi') **oida** ('tenho visto', 'sei') **eidos** ('forma, formato, figura, espécie')

latim:	**video** ('eu vejo') **visum** ('aparência, rosto') **visio** ('visão')
alemão:	**wissen** ('saber')
inglês:	**wise** ('sábio, sagaz, sensato') **wit** ('sagacidade, perspicácia')
russo:	**videt** ('ver')

6.9.2 *Dando o que falar*

O verbo dar é um dos mais usados em todas as línguas do mundo porque permite um sem-número de derivações de sentido, frases feitas e expressões idiomáticas. Basta verificar nos dicionários a quantidade de acepções que ele tem na nossa língua. No entanto, só o olhar do especialista consegue ver dar em usos ainda mais numerosos, dos quais os falantes em geral nem desconfiam.

Da raiz indo-europeia *do temos o sânscrito dadāiti, o albanês dashë, o lituano dúomi, o letão duôdu, o tcheco dán etc. Em grego temos o infinitivo dounai, do qual deriva a palavra dosis, origem do nosso dose — uma dose é algo que se dá a alguém (uma bebida, um remédio). Ainda do grego, herdamos ane.do.ta ('coisa inédita, ainda não dada à divulgação'), antí.do.to ('o que se dá contra [anti-]'), em que a raiz *do- está bem presente. Muitos nomes próprios ainda usados por nós têm origem nessa raiz, por intermédio do grego: Teo.doro, Deo.doro, Doro.teu ou Teo.dósio ('presente de Deus'), Eu.doro ('liberal, dadivoso'), Isa.dora ('presente de Ísis") etc. O nome Dora parece ser uma redução de Doroteia, feminino de Doroteu.

Em latim, temos o verbo dare (*do*, 'eu dou'), particípio passado *datu-*, um verbo que deu muito o que falar. Com o prevérbio *ad-*, o verbo *dare* sofreu apofonia [▸270], tornando-se *áddere*, fonte do nosso *adir* e seus derivados: *adição, adicionar, adido, aditar, aditivo*, onde a raiz mais antiga só se preservou num *d*. Nem precisa dizer que de *dare* provém o nosso *dar*, junto com *dado, dadeiro, dádiva, dador, dativo*, entre outros. Em latim também existia o verbo *donare*, com o sentido de 'dar de presente', sendo que o presente dado se dizia *donum*, de onde vem o nosso *dom*. Desse verbo, com a queda regular do *-n-* intervocálico [▸300], surgiram *doar, doação, doador* etc., e também *donatário*, assim como *per.do.ar* e seus muitos derivados. Um substantivo derivado de *do* é, já em latim, *dos, dotis*, que nos deu *dote, dotar, dotação* etc.

Com o prefixo *man-*, 'mão', acoplado ao verbo *dare*, surgiu man.dare, origem do nosso mandar que, acrescido do prevérbio co-, deu origem a comandar.

Com o prevérbio *com-* + *dare*, temos o verbo latino cóndere, de onde provém o nosso es.conder (do latim abs.cóndere) e também re.côndito e as formas eruditas absconder, absconso etc.

Com o prevérbio *ex-*, mais assimilação e apofonia, se formou em latim o verbo éddere, com particípio éditum, de onde vem o substantivo editio, editionis, origem do nosso edição e tudo o que lhe diz respeito: edital, editar, editável, édito, editor, editora, editoração, editoria, inédito etc. Aqui também, só escavando muito é que encontramos nessas palavras a raiz do-. O mesmo acontece com *per-* que, se juntando a *dare*, fez surgir o verbo pérdere, com o sentido de 'dar completamente' e, por extensão, 'gastar inutilmente' e, por fim, 'perder'. É a fonte do nosso perder e de sua grande família: perda, perdido, perdedor, perdição, perdulário, imperdível etc., além de es.perdi.çar e derivados e, é claro, des.perdi.çar e família.

Outro verbo em que ocorreu apofonia foi réddere, 'dar de volta, restituir, pagar' etc., formado de *red-* + *dare*. No latim vulgar, por influência de seu antônimo prénde-re ('pegar, tomar, apanhar'), o verbo *réddere* ganhou uma nasalidade, tornando-se réndere ('devolver'). É a fonte do nosso render e seus parentes: renda, rendição, rendimento, rendoso, arrendar, arrendatário e por aí vai.

Do particípio feminino data herdamos nossa data — abreviação de *charta data*, *epistula data*, *littera data*, fórmula usada nas cartas para indicar lugar e momento da escrita: *littera data in...* 'carta dada em...'.

Um verbo latino que, na sua derivação românica ocidental, não conservou nem sequer o *d* da raiz primitiva é trádere, formado de *trans-* + *dare*, com o sentido de 'entregar, transmitir, ceder, abandonar' etc. Com a queda regular do *-d-* intervocálico [▶308], o verbo se tornou primeiro *traer* e, mais adiante, trair — observe que ainda hoje usamos *entregar* com o sentido de trair ('Ele entregou os companheiros' / 'Ele traiu os companheiros'). A partir do século XVI, período da relatinização [▶143], surgiram novos termos em que o radical trad- foi recuperado: tradição, tradicional, tradicionalismo etc. Uma tradição é aquilo que se entrega, que se transmite às gerações futuras.

Uma história bonita parece se esconder por trás de outro derivado do verbo dare. É o verbo credo, infinitivo crédere, particípio passando créditum. Ao que parece, o primeiro elemento desse verbo, cre-, é uma mutação de cor, 'coração': *cor.do, credo. É o étimo do nosso verbo crer. O primeiro sentido de credo no latim clássico era 'emprestar', 'confiar alguma coisa a alguém' — ou seja 'dar de coração'. De crédere temos, inicialmente, creer (forma que se conservou em espanhol) e, após crase, crer. O radical cred- se conserva em muitas palavras nossas como credenciar, crédito, credor, crédulo, a.cred.itar etc. Já com radical popular temos crença, crendice etc. Os adjetivos crível e incrível também guardam, escondido, o velho radical latino.

Por fim, e não menos curioso, temos o verbo **vender**, derivado do latim **véndere**, formado da junção da palavra *venum* ('venda') com o verbo *dare*. O significado, portanto, é 'dar à venda'.

6.9.3 Espelho, espelho meu!

Da raiz indo-europeia *spek-, 'observar', temos uma rica derivação. Em latim existia o verbo **spēcĕre** (particípio passado: spectu-) que deu panos para manga em nossa língua: **aspecto, circunspecto, conspícuo, despiciendo, especial, espécie, específico, espe(c)táculo, espectador, espectar, espectro, espectral, espéculo, especular, especulação, expectação** (*ex-* + *spec-*), **expectativa, extrospecção, inspe(c)ção, inspe(c)tor, inspecionar, introspecção, introspectivo, perspectiva, perspicácia, prospecção, respectivo, respício, retrospecto, subespécie, subespecífico, suspicácia** etc. Nem é preciso dizer que de cada uma dessas palavras derivam outras tantas.

Mas a coisa não para por aí. Vamos ver em [▶312] que o grupo consonantal latino **-ct-** se transformou em **-it-** no galego arcaico. Essa mudança atingiu a raiz spect-, dando origem ao nosso radical **speit-**, que encontramos em **despeito, desrespeito, respeito, suspeito** e nas muitas palavras derivadas desses nomes.

Também veremos em [▶298] que o grupo latino **-cl-** se transformou em **-lh-** [ʎ]. Assim, a palavra latina speculu-, após ter perdido sua vogal postônica (spec'lu-) deu origem ao nosso **espelho** e à sua numerosa família.

Em latim também havia o elemento de composição **-spex, -spicis**, com que se formaram as palavras **au.spício** e **fronti.spício**. A palavra **auspício** provém de **auis** ('ave') e **-spiciu-**. De onde vem essa ave? Os romanos eram muito supersticiosos e tentavam fazer previsões e adivinhações examinando tudo o que estivesse à sua volta. Assim, o **áuspice** (do latim **auspex**) era aquele que fazia previsões observando o voo, o canto e o modo de comer das aves. As previsões favoráveis, é claro, eram as esperadas e por isso, até hoje, dizemos que algo é **auspicioso** quando traz vantagens e benefícios, embora nem de longe nos ocorra pensar em cegonhas, pardais e periquitos.

No grego, a raiz ***spek-** sofreu metátese [▶297] e se transformou em **skop-**, donde o verbo **skopeo** ('olhar atentamente, observar, examinar') e o substantivo **skopós** ('o observador atento'). É daí que vêm todas as nossas palavras (principalmente médicas) terminadas em **-scopia** (**endoscopia, uroscopia, radioscopia** etc.) e os nomes de aparelhos terminados em **-scópio** (**estetoscópio, microscópio, telescópio** etc.). Daí também provém a palavra **escopo**, 'alvo, ponto de mira' e depois 'intenção, objetivo' [▶839].

Por fim, do derivado germânico da raiz indo-europeia temos ***spehōn**, origem do nosso verbo **espiar** e toda a sua família.

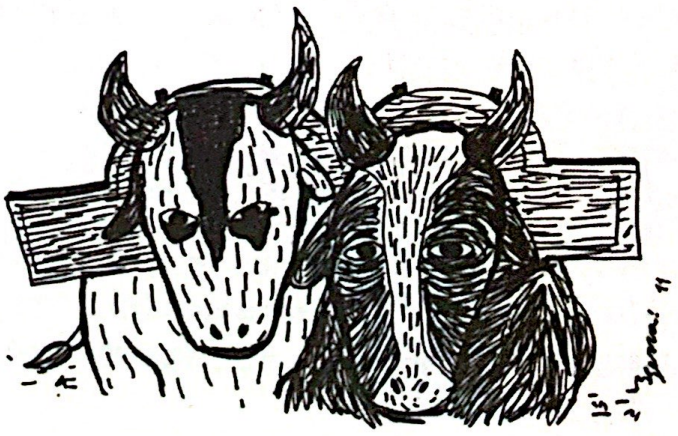
6.9.4 Todos juntos agora!

A raiz indo-europeia *yeug- tem o sentido de 'unir, juntar'. Em sânscrito ela gerou, entre outras, as palavras yugá ('jugo, parelha') e yogah ('união'), origem da nossa ioga. Em inglês, yoke ('jugo'). Em grego, zygón ('jugo'), o verbo zygóo ('unir'), que serviu de base para uma série de palavras técnicas e científicas como zigoma, zígio, zigócero, zigoto, sizígia ('conjunção ou oposição da lua com o sol') etc. Do grego também nos veio zeugma, uma figura de linguagem correspondente à elipse.

No latim, temos a palavra jugu- ('jugo') e o verbo jungĕre (particípio passado: junctu-) onde se verifica a ocorrência do infixo nasal, que vimos mais acima [6.8]. É daí que procedem as nossas palavras abjugar, abjunção abjungir, abjuntivo, abjunto, adjunção, adjungir, adjunto, ajumentação, ajumentado, ajumentar, ajunta, ajuntação, ajuntar, biga (de *bi-* + *juga*), bigário, bijugado, bíjugo, conjucida, conjucídio, conjugação, conjugal, conjugar, cônjuge, conjunção, conjungir, conjuntar, conjuntável, conjuntiva, conjuntivite, conjuntivo, conjunto, conjuntura, dejungir/desjungir, desajuntar, desconjugar, desconjuntação, desconjuntado, desconjuntar, desconjunto, desjungir, desjuntar, disjunção, disjungir, disjuntar, disjuntiva, disjunto, disjuntor, enjugamento, enjugar, esconjuntar, inconjugável, injunção, injungir, injuntivo, jugo, jugular, júgulo, jumenta (de *jugmenta*), jumentada, jumentice, jumento, junção, jungido, jungidor, jungimento, jungir, junta, juntamento, juntar, junto, juntura, rejuntamento, rejuntar, subjugação, subjugar, subjunção, subjuntivo etc.

Existia também em latim o advérbio juxta (de jug.sta), que serve de elemento de composição para palavras nossas como justapor, justaposição.

A palavra cônjuge tem uma história interessante. Quando as pessoas se casavam, na Roma antiga, recebiam de presente uma parelha de bois reunidos por aquilo que hoje chamamos de canga, uma peça de madeira que prende os dois animais. A canga em latim se chamava jugum, origem da nossa palavra jugo. Assim, cada cônjuge recebia esse nome porque compartilhava o mesmo jugo, isto é, a mesma parelha de bois, que simbolizava a união do casal. Como já vimos ao estudar a metáfora e sua incidência



sobre a mudança linguística [▶170], o termo concreto jugo teve seu sentido estendido para significar aquilo que oprime, que prende, origem do verbo subjugar.

6.9.5 A *família do rei*

Outra raiz indo-europeia interessantíssima é *reg-, com o sentido de 'mover em linha reta, conduzir'. Nas diferentes línguas da família, essa raiz gerou palavras bem importantes. Em sânscrito, a língua da vastíssima literatura religiosa da Índia, temos o verbo raji-, 'guiar', rāj, 'rei', (origem do nosso *rajá*) e rājni, 'rainha' (compare com o latim regina, 'rainha'). No irlandês antigo, rīg, 'rei', e rīgain, 'rainha'. Em galês, rhaith; em gótico, raíths; em antigo alto alemão, reht (alemão moderno, recht): 'reto'. Em latim rēx, 'rei', rēgīna, 'rainha'. Em gótico, reiks, 'rico, poderoso, forte' (origem do nosso adjetivo *rico*, presente nos nomes próprios que começam com *Ric-* e terminam com *-rico*, *-rigo* [▶215]). Está também na origem do inglês right ('reto', 'certo') e do alemão recht, com os mesmos sentidos. Em alemão também existe o elemento de composição richt-, presente em richtig ('certo', 'correto') e em Richtung ('direção', 'orientação'). Dessa mesma raiz, ainda em alemão, provém o termo Reich ('reino', 'império'). Até hoje, por exemplo, a França é chamada pelos alemães de Frankreich, 'reino franco', e a Áustria de Österreich, 'reino do leste'. Quando Hitler, na sua insanidade, quis criar um III Reich ('terceiro império'), o número III tentava fazer uma ligação histórica com o Sacro Império Romano Germânico, que vigorou a partir do século X, e que seria o "I Reich", e com o Império Alemão fundado em 1871, que seria o "II Reich".

Dessa raiz deriva o verbo latino rēgĕre, 'dirigir, governar', com o particípio passado rectu-, origem do nosso verbo *reger* e do adjetivo *reto* (que já foi escrito *recto*). Com esse verbo se formaram as palavras reitor (< *rectore-*), regência, regime, regimento, região, dirigir (com apofonia), endereçar (< *indirectiare*), corrigir, corre(c)to, direito (< *derecto*, latim vulgar), erguer (< *erigere*), ere(c)to, erigir, surgir (> *surgo* > *sub-rigo*), surto, ressurgir (particípio passado latino: *ressurectu-*, de onde ressurreição etc.), insurgente, insurre(c)to.

De rēx, rēgis temos rei, real (< *regale-*), realengo, realeza, regalia, regicida, régio, reino, reinar, reiúno, interregno etc.

Da mesma raiz temos em latim rēgŭla, de onde vêm nossas palavras régua, regra, relha, bem como regular, irregular, regulamento, arreglar etc.

Com alteração da vogal (e > o) surgiu em latim o verbo rogare, com o sentido primitivo de 'dirigir-se a'. Pelas extensões de sentido, o verbo passou a significar 'fazer uma pergunta a; interrogar; pedir (perguntando)'; em seguida, 'pedir, solicitar'; mais adiante, 'consultar'; logo, 'propor uma lei'; por fim, 'chamar, convocar'. Nosso verbo rogar significa, hoje, 'pedir insistentemente, instar, exortar'. Seus deri-

vados mais usuais são ab-rogar, ad-rogar, arrogância, derrogar, interrogar, ob-rogar, prorrogar, sub-rogar.

6.9.6 *Nosso código genético*

A raiz indo-europeia *gen-/gne-, 'gerar, nascer', deu tantos e tão diversos frutos que é impossível não se admirar com a sua riqueza.

Em grego, essa raiz se tornou génos, 'raça, origem, nascimento'. Dessa palavra herdamos uma série de termos técnicos e científicos: genética, alógeno, alotígeno, autógeno, biogênese, cianógeno, colágeno, estrógeno, eugenesia, eugenia, felógeno, fotogenia (donde fotogênico), genealogia, genearca, heterogêneo, homogêneo, hidrogênio, oxigênio, patogênico. Também do grego temos gênese e todos os seus derivados: androgênese, anfigênese, antogênese, antropogênese, autogênese, biogênese, cosmogênese, embriogênese, epigênese, filogênese, fitogênese, geogênese, glicogênese, heterogênese, hipergênese, homogênese, metagênese, ontogênese, partenogênese, patogênese, pedogênese, singênese, tanatogênese, termogênese e por aí vai.

Da palavra grega surgiu gene, vocábulo científico criado no século XIX, assim como genoblasto, genocídio, genócito, genocracia, genocrático, genoma, genotípico, genótipo etc. Temos também o sufixo -genia ('origem, causa'), empregada em inúmeras palavras científicas (eugenia, fisiogenia, zoogenia etc.), assim como -geno em endógeno, exógeno etc.

Ainda em grego, a raiz indo-europeia assumiu a forma gónos, que se tornou um elemento de composição para a formação de palavras científicas como gonorreia, entre outras. É também daí que vem o sufixo -gonia, usado em muitas palavras como, por exemplo, cosmogonia ('origem do universo'), teogonia ('nascimento dos deuses'), além de muitos termos científicos.

Em inglês, a raiz assumiu a forma kind, 'raça, estirpe'. Essa palavra também significa, como adjetivo, 'gentil'. Ora, a palavra gentil nos vem do latim e nela está presente a mesma raiz de gens, 'raça, estirpe'...

Continuando com o latim, temos a palavra genus (genitivo: generis), 'linhagem, origem', de onde nos vêm palavras como congênere, congenial, congênito, degenerar, general, generalizar, genérico, gênero, generoso, genial, gênio, genioso, genital, genitor, genuíno, ingênito, ingênuo, primogênito, progênie, progênito, progenitor, progenitura, regenerar entre centenas de outras.

A forma latina do acusativo era génerem. Com a queda regular do *-n-* intervocálico [▶300], temos primeiro geer- e em seguida, com a crase, ger-. Essa é a forma da qual se originam geral, gerar, geração, geralista, gerativo etc. Observe que geral e

general são formas divergentes [▸154] de um mesmo vocábulo latino, *generale-*. Na língua medieval ainda se escrevia *geeral* e *geerar*.

O radical latino gener- tinha acentuação tônica na primeira sílaba (géner-). Sabemos que um sem-número de palavras latinas proparoxítonas se transformaram em paroxítonas no galego [▸154]. Ora, o mesmo se deu com as palavras formadas a partir de géner-, sobretudo as de origem popular. Por exemplo, *géneru-*, além do vocábulo erudito gênero, também é o étimo de genro ('marido da filha'). Esse é um caso raro de preservação do encontro *-nr-*, que ocorre em pouquíssimas palavras da língua (*tenro, Henrique, enricar*). O mais comum é o surgimento de uma consoante [d], prolongamento do caráter dental do [n], mais apta a formar um encontro de articulação mais fácil com [r]. Isso explica, por exemplo, que o latim *téneru-* tenha se transformado no francês *tendre* ('terno, macio', após uma fase *tenru-*), assim como *géneru-* se transformou em *gendre* ('genro'). O mesmo ocorreu em nossa língua, onde encontramos a evolução géner- > genr- > gendr-. É esse último radical que está presente em engendrar e derivados.

Da palavra latina in.geniu- surgiu, com sândi interno [▸300], nosso radical engenh-: engenho, engenhoso, engenhar, engenharia, engenhoca etc.

Outra forma assumida pela raiz em latim é -gn-, presente em praegnante-, de onde provém o radical pregn-, que encontramos em pregnar, pregnância, impregnar e família. Vamos ver que o grupo latino *-gn-* se transformou em *-nh-* [▸312], de modo que o radical pregn-, na derivação por via popular, mudou em prenh-, origem de prenha, emprenhar, prenhe, prenhez etc. A mesma forma -gn- está presente em benigno e maligno.

Do latim gens (acusativo: gentem) vem nossa palavra gente — curiosamente, a palavra mais frequente no poema épico *Os Lusíadas*, de Camões que, não por acaso, canta as glórias e conquistas da *gente* portuguesa... É fácil listar os seus derivados: gentaça, gentalha, gentama, gentarada, gente, gentes, gentiada, gentiense, gentil, gentileza, gentílico, gentinha, gentio etc.

Também diretamente do latim nos veio indígena. Se alguém imagina que essa palavra tem qualquer coisa a ver com "índio", pode esquecer essa ideia. Ela é forma de indu- ('movimento para dentro') e -genu ('origem, nascimento'). O indígena, portanto, é aquele que 'nasceu dentro' de um lugar, de um território etc. Assim, uma pessoa nascida e criada na França e de secular ascendência francesa é um *indígena* daquele país.

Se você acha que já é o bastante, segure o fôlego, porque ainda tem mais. A raiz indo-europeia *gen- acrescida do sufixo latino -men fez surgir gen-men. Por efeito de dissimilação [▸152], formou-se a palavra germen (genitivo: germinis), 'germe, rebento, broto'. Já sabe o que *germinou* daí, não é? Germe/gérmen, germinação, germinado, germinar, germicida etc.

E daí também, para surpresa de muita gente, a palavra irmão... Como assim? Fácil: o étimo de irmão é o latim germanu-. Em sintagmas como *o germano, meu germano*, o [g] se encontrava em posição intervocálica e, como ocorreu regularmente com as consoantes sonoras intervocálicas, ele sofreu síncope [▶154]. Daí resultou a forma ermanu-, que evoluiu para ermão e em seguida para irmão. Quem vê a palavra gene e a palavra irmão dificilmente imaginará que são palavras... irmãs. Mas, afinal, chamamos de irmã(o) a pessoa que tem o mesmo germe que nós, não é?

Mais difícil talvez seja imaginar que gene e nação procedem de uma mesma raiz. Mas pode acreditar: é isso mesmo. A raiz *gna- se juntou ao sufixo -sko, gerando o verbo latino *gnascor. Ora, ao contrário do grego, em que as palavras podem começar com muitas combinações de consoantes (*tmese, ptose, ctônico, psíquico, pneu, gnose* etc.), os falantes do latim não se davam tão bem assim com elas. Desse modo, o *g* inicial sofreu apócope [▶296] e, no latim clássico, o verbo já apresentava a forma nascor (particípio passado: natu-). O mesmo aconteceu com o verbo *gnosco ('conhecer, saber') que se transformou em nosco — no entanto, o *g* reaparece em formas com *ignorar* e derivados, e também em *ignóbil*, que é o antônimo de *nobre* (< *nobile-*).

Desses radicais nasc- e nat- (nac- / nad-) vão surgir muitas e muitas palavras essenciais do nosso léxico (e do léxico de muitas línguas também): adnascente agnação, agnado, agnatício, cognação, cognato, co-nacional, conato, conatural, desnacionalizar, desnaturado, desnaturalizado, desnaturalizar, desnaturalizável, desnaturamento, desnaturar, desnaturável, enação, inascibilidade, inascível, inatismo, inatista, inato, inatural, inaturalidade, inaturável, internacional, internacionalizar, multinacional, nação, nacional, nacionalidade, nacionalismo, nacionalista, nacionalizar, nacionalizável, nadível, nadivo, nado, nascediço, nascedio, nascedoiro/nascedouro, nascença, nascente, nascer, nascida, nascidiço, nascido, nascimento, nascituro, natal, natalício, natalidade, natalino, natimorto, natividade, nativismo, nativista, nativitário, nativo, nato, natura, naturado, natural, naturaleza, naturalidade, naturalismo, naturalista, naturalizar, naturalizável, naturalogia/naturologia, naturalógico/naturológico, naturença, naturante, naturar, natureza, naturismo, naturologia, naturopatia, naturopático, naturoterapeuta, naturoterapêutico, prenascer, prenascido, pré-natal, prenato, preternatural, preternaturalidade, renascença, renascense, renascente, renascentismo, renascentista, renascer, renascido, renascimento, renacionalização, renacionalizado, renacionalizar, Renato, renaturar, sobrenatural, supernacional, supernatalidade, supernatural, supernaturalismo, transnacional, transnacionalizar etc.

Da locução *res nata* ('coisa nascida') é que provém o nosso nada. Quando, no latim vulgar, se dizia algo semelhante a "não sei *coisa nascida* a respeito disso", o que se estava dizendo era "não sei coisa nenhuma a respeito disso". Pelo processo de gramaticalização [▶170], o adjetivo nata, por metonímia, incorporou o sentido negativo

presente no enunciado completo, dispensando o substantivo *res* ('coisa'). Com a lenição [▶137] do [t] intervocálico, nata se transformou em nada. Esse processo também ocorreu em francês: quando aprendemos que a palavra personne significa 'ninguém' em francês, achamos estranho, já que pessoa, para nós, quer dizer justamente o contrário: 'alguém', 'algum indivíduo'. Aqui também ocorreu uma metonímia: na expressão *je ne vois personne* ('eu não vejo pessoa'), o substantivo *personne* incorporou o valor negativo de *ne*. É por isso que, hoje em dia, se alguém perguntar em francês: "*Qui t'a aidé à faire ça?*" ('Quem te ajudou a fazer isso?"), é possível responder: *Personne!* ('Ninguém!').

Finalmente, do latim cognatu- temos a forma cunhado e seus derivados.

6.9.7 O selo, o sino e a senha

O substantivo latino signum está na origem de uma grande quantidade de palavras do nosso léxico. Na sua forma erudita, mais próxima do latim, portanto, a raiz nos deu termos como signo, significado, significante, significar, consignar, designar, desígnio, insígne, insígnia, insignificante, persignar, resignar etc.

Na sua forma semierudita, a raiz sign- perdeu o -g- e se tornou sin-, presente em assinalar, assinar, ensinar, sina, sinal, sineta, sino e por aí vai.

Na sua forma popular, a raiz assumiu a forma senh-, que podemos ver em diversas formas divergentes [▶154], em que um mesmo étimo latino deu origem a duas palavras nossas: designiu- > desígnio/desenho; resignare > resignar/resenhar; signu- > signo/senho. Do plural neutro signa provém a nossa senha, tão necessária hoje para termos acesso aos serviços do mundo virtual.

Por fim, de sig.illum, diminutivo de sig.num, temos sigilo e seus derivados. Do mesmo sigillum, pelas mudanças regulares, nos vem a palavra selo que, inicialmente, designava o lacre de cera que preservava o sigilo dos documentos oficiais.

LIVRO III MULTIMÍDIA DO PORTUGUÊS BRASILEIRO

a Antônio Augusto Souza Mello

7

os sons e os séculos

fonologia da nossa língua

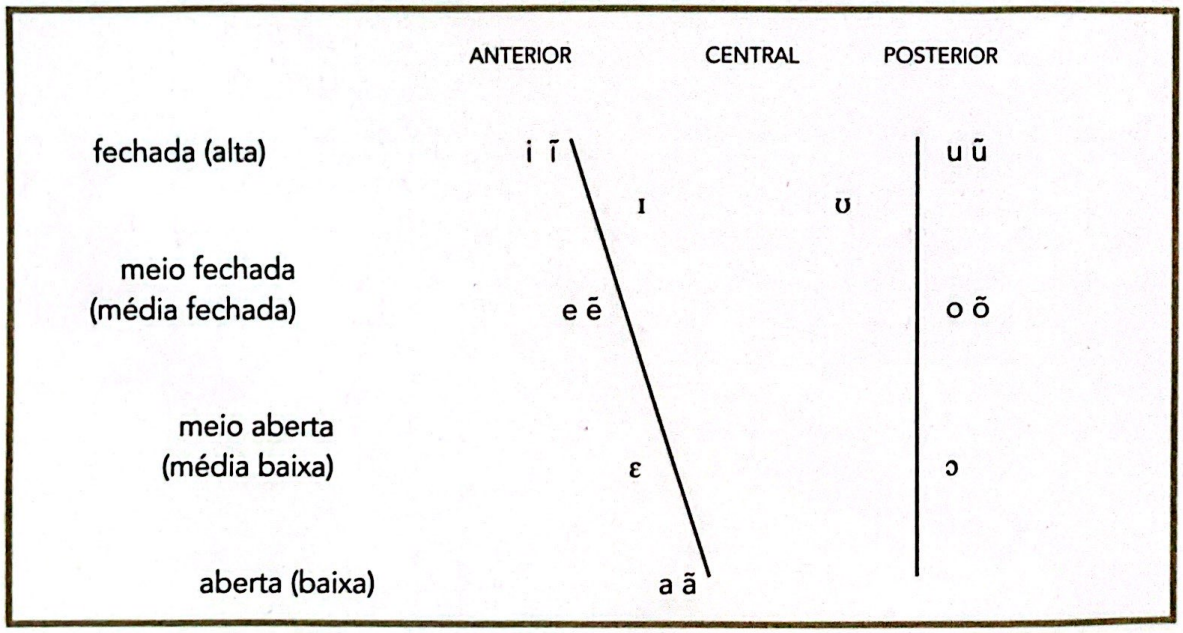
O latim que foi levado para as diferentes províncias do Império Romano era um conjunto de variedades que a linguística histórica classifica de latim vulgar. Depois da dissolução do Império, os diferentes reinos novos, instaurados e governados pelos conquistadores germânicos, ficaram de certo modo isolados uns dos outros. Com isso, sem haver

mais a força centrípeta das instituições romanas, o latim vulgar foi se transformado segundo ritmos e tendências distintas em cada uma das regiões romanizadas. Nessa transformação, o papel dos substratos e dos superestratos, como vimos, foi sem dúvida importante.

Nesse capítulo, vamos examinar as mais notáveis mudanças fonéticas ocorridas no noroeste da Península Ibérica, a Galécia, berço da língua galega, língua que mais tarde, sendo levada cada vez mais para o sul, se tornará o português europeu. Em seguida, vamos ver como esses sons viajaram no tempo e no espaço até chegar no Brasil e formar o PB que falamos hoje.

7.1 Vogais e consoantes

Para um bom acompanhamento da história das mudanças fonéticas, é importante conhecer o quadro dos sons do PB. Vamos iniciar pelas vogais:



É sempre útil recordar que esse desenho representa o chamado espaço fonológico, ou seja, em termos simples, o interior da nossa boca. Quando classificamos uma vogal como baixa é porque ela se produz mais perto da língua, na parte baixa da boca. E quando é alta, ela se produz mais perto do palato (o céu da boca), no alto da cavidade bucal. Por isso também as designações aberta e fechada se referem ao tanto que abrimos ou fechamos nossa boca para a produção das vogais. Assim, o [a] é a vogal mais aberta, pois exige uma abertura total da boca. À medida que vamos fechando a boca, as vogais também vão se fechando. Todas as vogais são, na verdade, o som [a] emitido com maior ou menor fechamento da boca. As vogais chamadas

posteriores são todas arredondadas, ou seja, para serem produzidas exigem que nossos lábios se projetem (fazendo um "bico") e vão se fechando aos poucos até o grau máximo, que é [u]. Quanto mais alta é uma vogal anterior, a sua produção se realiza mais próxima do palato. A vogal mais alta anterior é o [i], e sua proximidade do palato vai ser responsável por uma série de fenômenos de mudança fonética na história da língua. Uma delas, específica do PB, são as pronúncias [ʧ] e [ʤ] das consoantes [t] e [d] quando seguidas de um [i] — essa vogal produz a palatalização das consoantes em palavras como *tia* e *dia* na maior parte das variedades do PB.

Para o estudo das mudanças fonéticas é sempre conveniente ter esse desenho em mente, porque, observando a posição das vogais, fica mais fácil compreender alguns processos ocorridos na história da língua. Por exemplo, o ditongo latino au se transformou, no galego, no ditongo ou, por força da assimilação: o [ʊ] muito alto, atraiu para mais perto de si a vogal mais baixa, [a], fazendo ela se fechar em [o]. Numa etapa seguinte, no português europeu e no PB, o ditongo escrito ou se monotongou em [o] concluindo o processo de assimilação, que não foi acompanhado pela ortografia (nem pelo galego, que conserva até hoje o ditongo [oʊ]). O mesmo se deu em castelhano e em toscano, línguas em que a ortografia seguiu a mudança, de modo que nessas línguas se escreve oro, tesoro, poco, enquanto nossa ortografia ainda registra ouro, tesouro, pouco. Mais conservadora ainda, a ortografia do francês continua a escrever au para o que se pronuncia [o]: *pauvre* ('pobre'), *chaud* ('quente'), *faux* ('falso'). As relações entre o que se fala e o que se escreve, portanto, variam muito de uma língua para a outra, a depender de circunstâncias que não são linguísticas, mas sim políticas, culturais, ideológicas. Tem muito mais de política e ideologia nas normas ortográficas do que uma visão ingênua poderia supor.

Vamos examinar agora o quadro das consoantes da nossa língua:

MODELO DE ARTICULAÇÃO		PONTO DE ARTICULAÇÃO						
		bilabial	labio--dental	dental ou alveolar	alveo--palatal	palatal	velar	glotal
oclusiva	desv.	p		t		c	k	
	voz.	b		d		ɟ	g	
africada	desv.				ʧ			
	voz.				ʤ			
fricativa	desv.		f	s	ʃ		X	h
	voz.		v	z	ʒ		ɣ	ɦ
nasal	voz.	m		n		ɲ		
tepe	voz.			ɾ				
vibrante	voz.			r̆				
retroflexa	voz.			ɹ				
lateral	voz.			l (ɫ)		ʎ		

[Baseado em Cristófaro Silva (1999), com adaptações.]

No quadro anterior, a abreviação *voz.* significa "vozeada", isto é, uma consoante produzida com a participação das pregas vogais (em estudos tradicionais, se usa a designação sonora), enquanto *desv.* significa "desvozeada", sem a participação das pregas vocais (surda). Para acompanhar bem as transformações ocorridas na língua, é importante observar que existem pares de consoantes em que a única diferença é precisamente o traço do vozeamento: [p] e [b], [t] e [d], [k] e [g], [f] e [v], [s] e [z], [ʃ] e [ʒ]. Como vimos ao estudar os efeitos de substrato, as consoantes desvozeadas intervocálicas do latim vulgar se transformaram em suas homorgânicas vozeadas — homorgânica é a consoante de um par que se diferencia da outra por um traço distintivo, nesse caso a diferença entre ser vozeada (sonora) ou desvozeada (surda).

No quadro das consoantes, é fácil para um falante de PB depreender o som que a maioria delas representa, porque muitos símbolos fonéticos são idênticos às letras que usamos na escrita. Outros, no entanto, precisam de alguns esclarecimentos:

(a) [ʃ] e [ʒ] correspondem, respectivamente ao som consonantal presente em *chá* e em *já*. Em muitas variedades do PB, como sabemos, ocorrem as consoantes [ʧ] como em *tia* ("tchia") e [ʤ] como em *dia* ("djia") resultantes da palatalização de [t] e [d] provocada pela vogal alta [i].

(b) [ɲ] e [ʎ] correspondem, respectivamente, ao que grafamos como *nh* e *lh* em *tenha* e *telha*. Essas consoantes, que não existiam em latim, também são resultantes da palatalização provocada pela semivogal [ɪ], como veremos adiante.

(c) [ɬ] representa uma velarização de [l]. Na grande maioria das variedades do PB, a consoante [l] em final de sílaba (e de palavra) se vocalizou em [ʊ]. No entanto, em algumas variedades, sobretudo no Sul e em faixas etárias mais avançadas, a vocalização ainda não se concluiu, permanecendo no estágio velarizado [ɬ]. Essa consoante é a mesma que ocorre em inglês e em português europeu nas mesmas condições de final de sílaba (e de palavra).

(d) Os símbolos [c] e [ɟ] representam a palatalização das consoantes [k] e [g] respectivamente diante das vogais [ɛ], [e] e [i]. Se você prestar atenção ao modo como fala, perceberá que as sílabas [kɛ], [ke], [ki] e [gɛ], [ge], [gi] não são pronunciadas da mesma forma como pronunciamos [ka], [kɔ], [ko], [ku] e [ga], [gɔ], [go] e [gu]. Como as vogais daquelas primeiras sílabas são produzidas perto do palato, as consoantes [k] e [g], que são pronunciadas na região velar (onde está o véu palatino, também chamado de palato mole), são "puxadas" para adiante, adquirindo um som palatalizado. Foi por isso que, como vamos ver mais adiante, essas sílabas, na transformação do latim nas línguas românicas, passaram por muitas mudanças. Em latim clássico, por exemplo, o nome *Cícero* se pronunciava ['kikero]. Isso também explica por que as sílabas escritas *ce-* e *ci-* soam

como [sɛ], [se], [si], enquanto as que são escritas *ge-* e *gi-* soam como [ʒɛ], [ʒe] e [ʒi]. Entretanto, não vamos usar os símbolos [c] e [ɟ] quando transcrevermos foneticamente nossos exemplos, porque essa palatalização é generalizada entre os falantes das línguas do grupo portugalego e não servem, portanto, como indicadores de variação regional (ao contrário de [ʧ] e [ʤ]).

(e) A letra *r* e o dígrafo *rr* são símbolos gráficos empregados para representar uma série de pronúncias diferentes, de acordo com a região de origem do falante. Fica nítido, portanto, que é o som que apresenta maior variação no PB. Na tabela acima, aparecem nada menos do que sete símbolos para representar essas pronúncias, que vamos analisar mais abaixo [▶323].

7.2 Acento tônico

O acento tônico é a maior inflexão da voz com que a gente pronuncia determinada sílaba de uma palavra. Na transformação do latim vulgar em galego, a regra geral é a permanência da sílaba tônica.

No latim clássico, a posição do acento tônico dependia da quantidade das sílabas: não existiam palavras oxítonas; os dissílabos eram paroxítonos; os polissílabos tinham acento na penúltima sílaba se ela fosse longa (am<u>ā</u>tur) e na antepenúltima sílaba, se ela fosse breve (leg<u>ĭ</u>mus). No latim vulgar, perdeu-se a noção de quantidade, substituída pelo acento de intensidade. As vogais, que eram longas e breves, passaram a ser abertas e fechadas.

Dois fenômenos constantes estiveram presentes na passagem do acento do latim clássico para o vulgar:

(a) Nas palavras proparoxítonas em que a última sílaba tinha um encontro consonantal formado de uma oclusiva + [ɾ], o acento tônico sofria diástole (isto é, deslocamento do acento tônico para a sílaba posterior): *cáthedram* > *cathédra* > *cadeira*; *ténebras* > *tenébras* > *trevas*.

(b) Ocorria diástole sempre que havia um hiato com [i] tônico: *paríetem* > *pariétem* > *parede*; *mulíerem* > *muliére* > *mulher*.

7.3 Metaplasmos

Um metaplasmo é uma mudança na estrutura fonética de uma palavra, ocasionada por acréscimo, remoção ou deslocamento dos sons que a compõem. Na história da

nossa língua é possível localizar diversos metaplasmos que ocorreram com certa regularidade.

7.3.1 Metaplasmos por acréscimo

1. Prótese é o acréscimo de um segmento sonoro no *início* da palavra: *stare* > *estar*; *spiritu* > *espírito*; *scutu* > *escudo*.

   (a) Um caso especial de prótese é a aglutinação: incorporação do artigo no início da palavra: *lacuna* > *alagoa*; *minacia* > *ameaça*; *lesione* > *aleijão*.

2. Epêntese é o acréscimo de um segmento sonoro no *interior* da palavra: *stella* > *estrela*; *humile* > *humilde*; *úmeru* > *ombro*.

   (a) Uma modalidade particular de epêntese é o suarabácti (ou anaptixe), a intercalação de uma vogal para desfazer um grupo de consoantes: *planu* > *prão* > *porão*; *blatta* > *brata* > *barata*; *kruppa* (germânico) > *grupa* > *garupa*.

3. Paragoge (ou epítese) é o acréscimo de um segmento sonoro no *final* da palavra: *ante* > *antes*.

7.3.2 Metaplasmos por supressão

1. Aférese é a supressão de um segmento sonoro no início da palavra: *acume* > *gume*; *attonitu* > *tonto*; *glattire* > *latir*; *blastemare* > *lastimar*.

   (a) Caso especial de aférese é a deglutinação, supressão de um *a* ou *o* inicial por confusão com o artigo: *horologiu* > *orologiu* > *relógio*; *apotheca* > *abodega* > *bodega*; *episcopu* > *obispo* > *bispo*; *abbatina* > *batina*.

2. Síncope é a supressão de um segmento sonoro no *interior* da palavra: *legale* > *leal*; *legenda* > *lenda*; *malu* > *mau*. Ocorreram muitíssimos casos de síncope regular na história da língua, como vamos ver mais adiante.

   (a) Uma modalidade de síncope é a haplologia, supressão da primeira de duas sílabas sucessivas iniciadas pela mesma consoante: *bondade* + *-oso* = *bondadoso* > *bondoso*; *trágico* + *comédia* = *tragicocomédia* > *tragicomédia*; *formica* + *-cida* = *formicicida* > *formicida*; *dedo* + *duro* + *-ar* = *dedodurar* > *dedurar*.

3. Apócope é a supressão de um segmento sonoro no *fim* da palavra: *mare* > *mar*; *amat* > *ama*; *male* > *mal*.

4. Crase é a fusão de duas vogais iguais numa só: *pede* > *pee* > *pé*; *colore* > *coor* > *cor*; *nudu* > *nuu* > *nu*. É um recurso para a eliminação do hiato. Com a síncope

de muitas consoantes intervocálicas, surgiram hiatos que, mais adiante, foram eliminados por meio da crase.

5. Sinalefa (ou elisão) é a queda da vogal final de uma palavra, quando a palavra seguinte começa por vogal: *de + intro > dentro; de + ex + de > desde; outra + hora > outrora.*
   (a) A sinalefa é fenômeno extremamente comum na fala corrente, em que os elementos da cadeia falada se aglutinam: O *Pedr'é um car'alegr'inteligent'e generoso.* É um recurso usado na versificação para a contagem das sílabas métricas: *quando eu*, pronunciado *quand'eu*, tem duas sílabas métricas. Quando algumas dessas aglutinações se tornam regulares e se *gramaticalizam*, ocorre o surgimento de palavras novas: *outrora, embora, hidrelétrica, radiativo, psicanálise* etc.

7.3.3 Metaplasmos por transposição

Podem ocorrer por deslocamento de um segmento sonoro ou pelo deslocamento do acento tônico da palavra:

1. Metátese é a transposição de um segmento sonoro para outra posição na sílaba, ou de uma sílaba para outra posição na palavra: *pro > por; semper > sempre; inter > entre; guirlanda > grinalda; capio > caibo; primariu > primairu > primeiro; fenestra > feestra > fresta.*

2. Hiperbibasmo é o deslocamento do acento tônico.
   (a) Quando o acento recua para a sílaba anterior, tem-se a sístole: *pantânu > pântano; campâna > campa; idólu > ídolo; erámus > éramos; livellu > nível.*
   (b) Quando o acento recua para a sílaba posterior, tem-se a diástole: *límite > limite; océanu > oceano; gémitu > gemido; íntegru > inteiro.*

7.3.4 Metaplasmos por transformação

1. Vocalização é a transformação de uma consoante em vogal: *nocte > noite; regnu > reino; multu > muito; absentia > ausência; factu > feito; actu > auto; capsa > caixa.*

2. Consonantização é a transformação de uma vogal em consoante. Ocorreu amplamente na transformação das semivogais latinas *i* e *u* nas consoantes *j* e *v* das línguas românicas: *iam > já; ieiunu > jejum; Hieronymu > Jerônimo; Iesus > Jesus; uacca > vaca; uita > vida.*

3. Nasalização é a transformação de um segmento oral em nasal: *nec > nem; mihi > mi > mim; sic > sim.* Ocorreu nasalização em muitas palavras que em latim

apresentavam o grupo inicial *ex-*, transformado em *enx-* ou *ens-*: *exagiu* > *ensaio*; *exame* > *enxame*; *exiectare* > *enjeitar*; *exsucare* > *enxugar*.

4. Desnasalização é o fenômeno inverso, em que um segmento nasal passa a oral: *luna* > *lũa* > *lua*; *corona* > *corõa* > *coroa*; *persona* > *pessõa* > *pessoa*. Na formação do galego e do português, a consoante [n] intervocálica sofreu síncope regular, deixando no entanto seu traço nasal sobre a vogal anterior. Mais tarde, esse traço nasal desapareceu em algumas palavras e se conservou em outras, provocando o surgimento da consoante palatal [ɲ] [▶300].

5. Sonorização ou lenização (abrandamento) [▶137] é a transformação de uma consoante surda na consoante sonora homorgânica. As consoantes latinas [p, t, k, f, s], quando mediais intervocálicas, se sonorizaram regularmente em [b, d, g, v, z]: *lupu* > *lobo*; *uita* > *vida*; *caecu* > *cego*; *profectu* > *proveito*; *acutu* > *agudo*; *acetu* > *azedo*; *vicinu* : *vizinho*.

   (a) Também ocorreu o abrandamento [b] > [v] classificado de *degeneração*: *rabia* > *raiva*; *rubeu* > *ruivo*; *arbore* > *árvore*.

6. Palatização ou palatalização é a transformação de um ou mais segmentos numa consoante palatal. O latim não tinha consoantes palatais. As que existem nas línguas românicas são, portanto, resultantes dessas transformações. No caso do português, temos:

◗ [ne, ni] + vogal > [ɲ] (grafada NH):	*vinea* > *vinha* *aranea* > *aranha* *seniore* > *senhor* *iuniu* > *junho*
◗ [le, li] + vogal > [ʎ] (grafada LH):	*palea* > *palha* *folia* > *folha* *iuliu* > *julho* *filiu* > *filho*
◗ [de, di] + vogal > [ʒ] (grafada J):	*video* > *vejo* *hodie* > *hoje* *invidia* > *inveja* *adiutare* > *ajudar*
◗ [pl, kl, fl] > [ʧ] (grafada CH): o galego conserva a pronúncia [ʧ], que em português se simplificou em [ʃ]	*pluvia* > *chuva* *implere* > *encher* *clave* > *chave* *flamma* > *chama* *inflare* > *inchar* *plumbu* > *chumbo*
◗ [kl, pl, gl, tl] mediais > [ʎ] (grafada LH):	*oculu* > *oclu* > *olho* *apicula* > *apecla* > *abelha* *scopulu* > *scoplo* > *escolho* *tegula* > *tegla* > *telha* *vetulu* > *vetlu* > *velho*
◗ [ske, ski, se, si] > [ʃ] (grafada X):	*pisce* > *peixe* *passione* > *paixão* *miscere* > *mexer* *russeu* > *roxo*
◗ [si] + vogal > [ʒ] (grafada J):	*basium* > *beijo* *caseum* > *queijo* *cerevisia* > *cerveja* *ecclesia* > *igreja*

7. Assibilação é a transformação de um ou mais segmentos sonoros numa consoante sibilante: *capitia* > *cabeça*; *audio* > *ouço*; *judiciu* > *juízo*. Observe-se que todos esses casos de assibilação se deveram à presença de um [ɪ] subsequente à consoante que se assibilou. Na fase arcaica da língua, o *ç* soava [ts] e o *z*, [dz]. Só bem mais tarde essas consoantes perderiam sua oclusão.

8. Assimilação é a mudança de um segmento sonoro num segmento igual ou semelhante a outro existente na mesma palavra: *ipso* > *isso*. Fenômeno muito importante na história da nossa língua, a assimilação pode ser total, parcial, progressiva e regressiva.

   (a) A assimilação é *total* quando o som assimilado se iguala ao som assimilador:
   persona > *pessoa*
   persicu > *pêssego*
   mirabilia > *maravilha*
   per + lo > *pello* > *pelo*

   (b) É *parcial* quando o som assimilado apenas se assemelha ao assimilador:
   auru > *ouro*
   lacte > *laite* > *leite*
   paucu > *pouco*

   (c) É *progressiva* quando o som assimilador está antes do assimilado:
   amam-lo > *amam-no*

   (d) É *regressiva* quando o som assimilador vem depois do assimilado:
   captare > *cattar* > *catar*

9. Dissimilação é a diferenciação de um segmento sonoro, devida quase sempre à existência de outro igual ou semelhante na palavra: *liliu* > *lírio*; *memorare* > *membrar* > *lembrar*; *rotundo* > *rodondo* > *redondo*; *locusta* > *logosta* > *lagosta*; *colonello* > *coronel*.

   (a) A dissimilação pode, às vezes, resultar na supressão de um segmento (dissimilação eliminadora), quase sempre a líquida [ɾ]: *aratru* > *arado*; *cribru* > *crivo*; *rostru* > *rosto*.

   (b) Convém registrar que os casos de assimilação são incomparavelmente mais frequentes e regulares na história da língua do que os casos de dissimilação.

10. Apofonia (também conhecida por seu nome em alemão: *Ablaut*) é a mudança de timbre de uma vogal por influência de um prefixo: *in + aptu* > *inepto*; *in + barba* > *imberbe*; *sub + jactu* > *sujeito*. É fenômeno que remonta à formação da língua latina, não tendo ocorrido na transformação do latim em galego.

11. Metafonia (ou *Umlaut*) é a mudança do timbre de uma vogal por influência do timbre da vogal ou semivogal seguintes: *totu* > *tudo*; *feci* > *fizi* > *fiz*; *décima* > *dízima*.

   (a) A metafonia pode ocorrer no singular, no masculino e na 1ª pessoa do indicativo, enquanto no plural, no feminino e na 2ª e na não-pessoa do indicativo se conserva o timbre original latino:

jocu > *jogo*; *focu* > *fogo*, mas *jocos* > *jogos*; *focos* > *fogos*; *eu jogo*
porcu > *porco*; *soceru* > *sogro*, mas *porca* > *porca*; *socera* > *sogra*
texo > *teço*; *verto* > *verto*, mas *texis* > *teces*; *vertis* > *vertes*

12. Sândi é um fenômeno que consiste na modificação de um segmento sonoro por influência de um som vizinho na cadeia falada. Por exemplo, o artigo *os* apresenta, em sua forma básica, um [s] final, mas quando combinado, na sintaxe, com um vocábulo iniciado por vogal ou consoante sonora, esse [s] se transforma em [z]: *os carros* [us'kaxus], mas *os amores* [uza'moris]; *os três* [us'treɪs], mas os dois [uz'doɪs]. A terminação *-em* se pronuncia, no PB, [e͡ĩ]; na cadeia sintagmática, se a palavra seguinte começa por vogal, desenvolve-se, por sândi, uma consoante [ɲ]: *sem dor* [se͡ĩ'doh], mas *sem amor* [sẽɲa'moh]. Assim, a palavra *senhora* e o sintagma *sem hora* apresentam a mesma pronúncia (se o falante não fechar o [e] de *senhora* em [i] como ocorre em algumas variedades: [sĩ'ɲɔra]).

(a) O sândi que ocorre entre palavras na cadeia sintagmática é chamado *sândi externo*. São os casos que vimos acima. Quando ocorre no interior da palavra é chamado *sândi interno*. Esse fenômeno foi responsável, em parte, pelo surgimento, no galego e no português, da consoante palatal [ɲ] desconhecida em latim.

(b) O *-n-* intervocálico sofreu síncope regular, deixando sua ressonância nasal sobre a vogal anterior:
arena > *arẽa*
bona > *bõa*
camminu > *camĩo*
corona > *corõa*
luna > *lũa*
pinu > *pĩo*
senu > *sẽo*
vinu > *vĩo*

(c) Numa fase posterior, essa ressonância nasal desapareceu nas palavras com o hiato *o-a*: *bõa* > *boa*; *corõa* > *coroa*, com conservação do hiato (muito mais tarde, no PB, esse hiato vai ser rompido com a epêntese de uma semivogal [ʊ]: *boa* ['boʊa]).

(d) No caso do hiato *ẽ-a*, a nasalidade desapareceu, mas logo se desenvolveu, por epêntese, uma semivogal [ɪ]: *arẽa* > *area* > *areia*; *sẽo* > *seo* > *seio*, com eliminação do hiato.

(e) Quanto à vogal [ĩ], por ser nasal e palatal, o hiato foi rompido por meio de uma extensão da nasalidade e da palatalidade para a sílaba seguinte, dando origem à consoante [ɲ], que passou a ser escrita *nh*: *vĩo* > *vinho*; *pĩo* > *pinho*; *camĩo* > *caminho*. Como se vê, é um caso de *sândi interno*, equivalente ao externo que ocorre, por exemplo, em *vim hoje* [vĩ'ɲoʒi].

7.3.5 Quadro-síntese dos metaplasmos

<table>
<tr><th colspan="5">ACRÉSCIMO</th></tr>
<tr><td colspan="2">PRÓTESE
(início da palavra)
minacia > ameaça</td><td colspan="2">EPÊNTESE
(interior da palavra)
blatta > brata > barata</td><td>PARAGOGE
(final da palavra)
ante > antes</td></tr>
<tr><th colspan="5">SUPRESSÃO</th></tr>
<tr><td>AFÉRESE
(início da palavra)
glattire > latir</td><td>SÍNCOPE
(interior da palavra)
legere > leer > ler</td><td>APÓCOPE
(final da palavra)
mare > mar</td><td>CRASE
nudu > nuu > nu</td><td>SINALEFA
de um > dum</td></tr>
<tr><th colspan="5">TRANSPOSIÇÃO</th></tr>
<tr><td colspan="2">METÁTESE
inter > entre; merulu > merlo > melro</td><td colspan="3">HIPERBIBASMO
sístole
pantánu > pântano
diástole
límite > limite</td></tr>
<tr><th colspan="5">TRANSFORMAÇÃO (1)</th></tr>
<tr><td>VOCALIZAÇÃO
octu > oito
CONSONANTIZAÇÃO
uacca > vaca
ieiunu > jejum</td><td>NASALIZAÇÃO
sic > sim
DESNASALIZAÇÃO
lumen > lume</td><td>APOFONIA
in + barba > imberbe
METAFONIA
porcu > p<u>o</u>rco [o]
porcos > p<u>o</u>rcos [ɔ]</td><td colspan="2">ASSIMILAÇÃO
paucu > pouco
DISSIMILAÇÃO
locusta > logosta > lagosta
rostru > rosto</td></tr>
<tr><th colspan="5">TRANSFORMAÇÃO (2)</th></tr>
<tr><td>SONORIZAÇÃO
(lenização)
acutu > agudo</td><td>PALATALIZAÇÃO
pisce > peixe
vinea > vinha</td><td>ASSIBILAÇÃO
capitia > cabeça
iustitia > justeza</td><td colspan="2">SÂNDI
vinu > vĩo > vinho
pinu > pĩo > pinho</td></tr>
</table>

7.4 Vocalismo

O latim clássico tinha cinco vogais: [a], [e], [i], [o], [u]. Segundo a *quantidade*, elas podiam ser breves ou longas. A essas diferenças de quantidade estavam associadas diferenças de timbre:

ĕ ĭ ŏ ŭ — breves e abertas
ē ī ō ū — longas e fechadas

O [a], breve ou longo, tinha o mesmo timbre.

Para um falante de PB, pode ser difícil imaginar um "i" ou um "u" abertos, visto que são as vogais mais fechadas que temos na língua. No entanto, para quem já estudou inglês, é possível ter uma ideia da diferença entre *ĭ* e *ī* pensando na diferença entre *it* ('isso') e *eat* ('comer'), e entre *ŭ* e *ū* pensando na distinção entre *luck* ('sorte') e *look* ('olhar'). Essas distinções, fundamentais em inglês, representam dificuldades para os falantes de outras línguas que não as têm em seu sistema fonológico. Perceba que o [i] de *it* se aproxima da vogal fechada [e], o que explica por que o *ĭ* se transformou em [e].

No latim vulgar, desapareceu a oposição quantitativa, e as vogais passaram a se diferenciar apenas pelo timbre, isto é, abertas e fechadas. No entanto, em toda a România, o *i* aberto se confundiu com o *e* fechado, e o *u* aberto se confundiu com o

o fechado. Desse modo, as dez vogais do latim clássico (cinco breves e cinco longas) se reduziram no latim vulgar a sete, conforme o quadro abaixo:

LATIM CLÁSSICO	LATIM VULGAR
ă ā	a
ĕ	ɛ
ē ī	e
ī	i
ŏ	ɔ
ō ŭ	o
ū	u

O galego (e, consequentemente, o português) conservou o mesmo quadro de vogais do latim vulgar, ao contrário de outras línguas românicas, como o espanhol (que não tem as vogais abertas) e o francês (que apresenta um quadro mais amplo do que esse, com a inclusão de vogais médias centrais como [y], [ø] e [œ]). As outras línguas românicas também passaram por vários processos de ditongação que não afetaram o galego e o português:

NOVU (lat. vulgar):

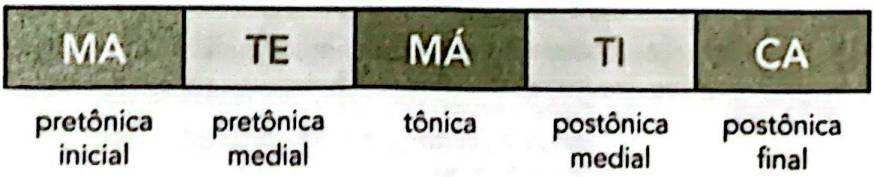
novo (galego)
nuevo (castelhano)
neuf (francês)
nou (catalão)
nuovo (italiano)
nouv (romanche)
nouvèu (provençal)
nou (romeno)

Para o estudo do vocalismo, é indispensável a distinção entre vogais tônicas e vogais átonas.

As vogais tônicas sofreram apenas as alterações já vistas no quadro acima. Elas se mantiveram em nossa língua em virtude de sua tonicidade:

LATIM CLÁSSICO	LATIM VULGAR	PORTUGUÊS
ăquam	a	água
pācem		paz
nĕbulam	ɛ	névoa
secrētum	e	segredo
cĭsta		cesta
rīvum	i	rio
rŏtam	ɔ	roda
sapōrem	o	sabor
bŭccam		boca
secūrum	u	seguro

As vogais átonas, quanto à sua posição na palavra, se classificam em pretônicas e postônicas, conforme estejam antes ou depois da sílaba postônica. As vogais pretônicas podem ser iniciais ou mediais, enquanto as postônicas são mediais ou finais:



No que diz respeito às vogais pretônicas:

(a) As pretônicas iniciais permaneceram ou sofreram aférese:

amicu > amigo — *acutu* > agudo
episcopu > bispo — *acume* > gume

(b) As pretônicas mediais sofreram síncope:

bonitate > bondade — *honorare* > honrar — *computare* > contar

Quanto às postônicas:

(a) As postônicas mediais adjacentes à tônica sofreram, em geral, síncope, devido à tendência dos falantes da Romania ocidental a evitar (até hoje) as proparoxítonas:

víride > verde — *lépore* > lebre — *ópera* > obra — *régula* > regra

(b) As postônicas finais *i* e *u* das palavras latinas passaram respectivamente a *e* e *o*: *vivi* > *vive*; *vesti* > *veste*; *lupu* > *lobo*; *libru* > *livro*. Na língua medieval, essas vogais finais eram pronunciadas [e] e [o] como até hoje em galego. No período clássico do português (séc. XV-XVI), elas sofreram alçamento e passaram a ser pronunciadas [ɪ] e [ʊ], pronúncia que é a normal do português brasileiro. No português europeu o *-e* final passou a ser pronunciado como uma vogal central média [ə] desconhecida no PB.

(c) A vogal *e* postônica final sofreu apócope quando precedida de consoante que podia formar sílaba com a vogal anterior (*l, n, r, s, z*): *male* > *mal*; *bene* > *bem*; *amare* > *amar*; *mense* > *mês*; *cruce* > *cruze* > *cruz*. A existência desse antigo *e* explica a forma do plural em *-es* de algumas dssas palavras: *meses, cruzes, males, luzes, flores*. Nas terminadas em *-l* a reinserção do *-e* no plural provocou a queda regular do *-l-* intervocálico: *anel* > *aneles* > *anees* > *anéis*; *funil* > *funiles* > *funiis* > *funis* etc.

(d) As vogais *a* e *o* postônicas finais não sofreram modificação: *aqua* > *água*; *Cicero* > *Cícero*. No período clássico, como vimos, o *o* átono final se transformou em [ʊ]. No português europeu, o *a* átono se transformou num [ɐ] mais fechado que o [a] tônico e que o [a] átono do PB.

1. Ditongos. O latim clássico apresentava quatro ditongos: *ae, au, oe* e *eu* (muito raro). No latim vulgar, já havia a tendência a reduzir esses ditongos a vogais simples: *celebs* (< *caelebs*), *sepis* (< *saepis*), *clostrum* (< *claustrum*).

Nas transformações que resultaram no galego e no português tivemos:

(a) *ae* pretônico → [i] ou [e]:	*aequale* > *igual*
	aetate > *idade*
	aestivo > *estio*
(b) *ae* tônico → [ɛ]:	*caelu* > *céu*
	caecu > *cego*
	quaero > *quero*
(c) *au* → [ou] e [au]:	*tauru* > *touro*
	paucu > *pouco*
	causa > *cousa*
	lauru > *louro*
	aula > *aula*
	audax > *audaz*
(d) *oe* → [e]:	*foedu* > *feo* > *feio*
	poena > *pena*
	foenu > *feno*

O galego e o português têm, portanto, dois ditongos de *origem latina*: [au] e [ou] (que passou a alternar com [oɪ] em fase mais recente do português: *cousa* ~ *coisa*). Convém lembrar que o ditongo [ou] tende, por assimilação, a ser monotongado em [o] tanto no português europeu quanto no brasileiro, em todas as posições em que ocorre e no vernáculo de todos os falantes, inclusive nas variedades urbanas mais prestigiadas. O galego, mais conservador nesse aspecto, conserva o ditongo [ou] até mesmo em palavras que soam arcaicas para os falantes de PB, como *cousa* e *dous*.

Além desses dois ditongos, nossa língua apresenta outros que são de *origem românica*, ou seja, formados na fase evolutiva dos romances. São diversas as condições que levaram à formação dos ditongos em português:

(a) síncope da consoante intervocálica:

vanitate > *vaidade*
malu > *mau*
palu > *pau*
vadi(t) > *vai*

(b) vocalização:

nocte > *noite*
regnu > *reino*
absentia > *ausência*
conceptu > *conceito*

(c) metátese:

primariu > **primairo* > *primeiro*
denariu > **dĭairo* > *dinheiro*
rabia > *raiva*
capio > *caibo*

(d) epêntese (de uma vogal para desfazer hiato):

cena > *cẽa* > *cea* > *ceia*
credo > *creo* > *creio*
frenu > *freẽo* > *freo* > *freio*

(e) oclusão (fechamento da segunda vogal do hiato, com desaparecimento do hiato):

malu > mao > mau
velo > veo > véu
amatis > amades > amaes > amais

O ditongo final *-ão* [ɐ̃ʊ̃] tão característico do português moderno, representa a transformação das formas arcaicas *-am, -ã, -om, -õ*, correspondentes às terminações latinas *-anu, -ane, -one, -ine, -unt, -um, -on, -ant, -a(d)unt*:

veranu > verão
pane > pão
oratione > oração
paganu > pagão
cane > cão
solitudine > solidão
dant > dão
sunt > são

Essas origens diversificadas explicam as diferenças na formação dos plurais dos nomes terminados em *-ão*:

manos > mã-os > mãos
leones > leõ-es > leões
panes > pã-es > pães

2. Hiatos. Desde a fase arcaica da língua, verifica-se uma notável tendência à eliminação dos hiatos. Os hiatos se desfizeram sob diferentes condições na história da língua:

(a) crase
- *vedere > veer > ver*
- *tenere > teer > ter*
- *magistre > maestre > meestre > mestre*
- *mala > maa > má*
- *populu > poboo > povo*

(b) oclusão
- *caelu > ceo > céu*
- *ego > eo > eu*
- *velu > veo > véu*

(c) epêntese
- *plenu > chẽo > cheo > cheio*
- *arena > arẽa > area > areia*
- *tela > tea > teia*
- *vinu > vĩo > vinho*

(d) desenvolvimento da palatal [ɲ]
- *mea > mĩa > minha*
- *nidu > nĩo > ninho*
- *pinu > pĩo > pinho*

7.5 Consonantismo

O latim clássico apresentava as seguintes consoantes:

articulação modo / lugar		bilabial	labiodental	dental ou alveolar	alveopalatal	palatal	velar	glotal
oclusiva	desv.	p		t			k	
	voz.	b		d			g	
africada	desv.							
	voz.							
fricativa	desv.		f	s				h
	voz.							
nasal	voz.	m		n				
tepe	voz.			ɾ				
vibrante	voz.			r				
retroflexa	voz.							
lateral	voz.			l				

Nesse quadro poderíamos incluir também as semivogais [ɪ] e [ʊ]. Comparando essa tabela com a das consoantes do PB vista mais acima, o nosso sistema fonológico é muito mais rico em consoantes do que o do latim clássico. Faltavam ao latim todas as consoantes das séries palatal e alveopalatal, além das fricativas sonoras [v] e [z]. Por outro lado, as consoantes latinas podiam ser breves ou longas, caso em que eram grafadas duplamente: *ager* ('campo') se distinguia de *agger* ('materiais amontoados, montão de terra'). Muitas dessas consoantes "geminadas" eram resultantes de processos de assimilação: *sub + ferre > subferre > sufferre*; *ad + captare > adceptare > acceptare*; *in + ludere > inludere > illudere* etc. Na formação dos romances, a distinção entre consoante simples e consoante dupla se perdeu, por ter pouco rendimento fonológico (isto é, por não ser fundamental na distinção de palavras de sentido diferente). Somente o italiano e o sardo conservam até hoje consoantes duplas em seus sistemas fonêmicos — cf. italiano: *note* ('notas') e *notte* ('noite').

No tocante à forma escrita, a letra V latina era simplesmente a forma maiúscula do *u* e não a consoante [v], que o latim não tinha. Por isso, em muitos monumentos de arquitetura neoclássica, é comum encontrarmos um V no lugar de um U (FORVM ou TEATRO MVNICIPAL).

Convém recordar que a aspiração indicada pela letra H em latim logo se perdeu, razão pela qual essa letra sobrevive hoje, nas línguas românicas, como mero indicador da etimologia da palavra. A letra C latina tinha sempre o som [k] (*Cícero* ['kikero]), e só na fase de formação dos romances ela sofreu influência das vogais [e] e [i] subsequentes e passou por diversas etapas fonológicas até se tornar a sibilante [s] que tal letra representa hoje em português diante de [e] e [i] como no nome próprio *Cecília*. O mesmo se diga da letra G do latim, que era sempre pronunciada [g] e passou por diversas transformações até se tornar, diante de [e] e [i] a palatal simples [ʒ] como na palavra *gengibre*. Isso ocorreu porque as consoantes [k] e [g] são

velares, isto é, pronunciadas junto ao véu do palato, na parte posterior do trato vocal, enquanto as vogais [e] e [i] são anteriores, pronunciadas nas proximidades do palato (céu da boca). Esse caráter palatal das vogais acabou influenciando a pronúncia das consoantes velares, de modo que o [k] se tornou a sibilante [s] (depois de passar por uma fase [ts]) e o [g], a palatal [ʒ] (depois de passar por uma fase [dʒ]).

Para representar sons da língua grega que não existiam no latim, os romanos criaram letras novas ou adaptaram algumas de suas letras:

(a) *ph, th, ch, rh* inicialmente não eram dígrafos: o *h* ali representado indicava uma aspiração que se seguia à primeira consoante, fenômeno fonético inexistente em latim (mas que existe no inglês quando as consoantes *p, t* e *k* iniciam palavra). Mais tarde é que a aspiração se fundiu com a oclusão (caso do *ph* que passou a ser pronunciado [f]) ou desapareceu (caso do *th* que passou a ser pronunciado simplesmente [t], do *ch*, pronunciado [k] e do *rh* pronunciado [ř]);

(b) o *y* foi usado pelos romanos para transcrever palavras gregas que tinham a vogal central alta arredondada que se representa em francês como *u* e em alemão como *ü* (símbolo fonético: [y]). Posteriormente, essa vogal passou a ser pronunciada como um simples [i]; em alemão, porém, até hoje nas palavras de origem grega o Y é pronunciado [y];

(c) o *z* representava a consoante grega ζ, pronunciada [dz]; na evolução do latim vulgar para as línguas românicas, essa letra passou a representar a sibilante sonora [z] (português, francês), permaneceu como [dz] em italiano (ou [ts] quando escrita geminada: *zz* (*pizza* ['pitsa]) e se tornou, em castelhano padrão e galego, a interdental [θ] (que a ortografia do inglês representa como *th* em *think*).

Como regra geral, as consoantes iniciais do latim se mantiveram inalteradas:

bene > bem	*gutta > gota*	*patria > pátria*
corona > coroa	*latrone > ladrão*	*rota > roda*
debere > dever	*manu > mão*	*salute > saúde*
filiu > filho	*navigiu > navio*	*tale > tal*

Há, no entanto, exceções: *cattu > gato; palore > bolor; libelu > nível; vesica > bexiga; vota > boda* etc. Algumas dessas mudanças já tinham ocorrido no latim; outras se devem a fatores diversos, como assimilação, analogia ou empréstimo linguístico.

Quanto às consoantes mediais, ocorreram os seguintes fenômenos:

(a) As consoantes mediais surdas, em posição intervocálica, se transformaram em suas homorgânicas sonoras:

p > b →	*ripa > riba*	*lupu > lobo*	*sapere > saber*
t > d →	*vita > vida*	*rota > roda*	*mutu > mudo*
k > g →	*focu > fogo*	*pacare > pagar*	*acutu > agudo*
f > v →	*profectu > proveito*	*trifoliu > trevo*	*aurifice > ourives*

s > z →	*acetu > azedo*	*vicinu > vizinho*	*placere > prazer*
	ca[s]a > ca[z]a	*ro[s]a > ro[z]a*	*pau[s]are > pou[z]ar*

(b) As consoantes mediais sonoras sofreram *síncope*, permaneceram ou se alteraram:

-b-	degenerou-se em [v]:	*caballu > cavalo*	
		faba > fava	
	ou sofreu síncope:	*ibam > ia*	
		praebenda > preenda > prenda	
-d-	sofreu síncope:	*sedere > seer > ser*	
		pede > pé	
		fidele > fiel	
-g-	sofreu síncope:	*regale > real*	*legere > leer > ler*
	vocalizou-se:	*plaga > praia*	*lege > lei*
	manteve-se:	rogare > *rogar*	*paganu > pagão*
-l-	*sofreu síncope:*	*filu > fio*	*velu > véu*
-m-	*manteve-se:*	*amicu > amigo*	*lacrima > lágrima*
-n-	nasalizou a vogal anterior e depois desapareceu como consoante. Na maioria dos casos, a ressonância nasal desapareceu: *persona > pessõa > pessoa; luna > lũa > lua; generale > gẽral > geral.* Nas terminações *-ina* e *-inu*, após a queda de *-n-* se desenvolveu a palatal [ɲ]: *vinu > vĩo> vinho; regina > reĩa > rainha.*		

(c) A síncope regular do *-l-* explica a forma dos artigos definidos em galego e português. Como os artigos *lo, la* ocorriam frequentemente entre vogais no encadeamento sintagmático, o *l-* inicial dessas palavras sofreu síncope também: *compro lo livro, quero la casa > compro o livro, quero a casa.*

(d) Ela também explica os plurais das palavras terminadas em *l*:

anel > aneles > anees > anéis
azul > azules > azues > azuis
canal > canales > canaes > canais
sol > soles > soes > sóis

Em geral, as consoantes finais latinas sofreram apócope. Só se conservaram em português *-m, -n, -r, -s*:

-m:	conservou-se nos monossílabos como simples ressonância nasal: *quem > quem* [kẽı]; *cum > com* [kõ].
-n:	permaneceu como ressonância nasal, ora representada por til, ora por *m*: *in > en > em* [ẽı]; *non > não.*

-r:	permaneceu, mas passou para antes da vogal precedente, por metátese: *inter > entre; super > sobre; semper > sempre; quattuor > quator > quatro.*
-s:	permaneceu como marca de plural (*lupos > lobos*); em alguns nomes próprios (*Deus > Deus; Marcus > Marcos*); nos advérbios (*magis > mais*) e nas desinências verbais: *amas > amas; amamus > amamos; amatis > amais* etc.

Pela queda da vogal final *-e*, as consoantes *-l-*, *-r-* e *-s-* mediais se tornaram finais: *fidele > fiel; legale > legal; mare > mar; debere > dever; mense > mês; portucalense > português; reverse > revés.*

O *-z-* medial, decorrente da sonorização do *-c-* intervocálico latino, se tornou final pela mesma razão: *luce > *luze > luz; radice > *raize > raiz; voce > *voze > voz; pace > *paze > paz.* A existência de um *-e* final na língua antiga explica a formação do plural dessas palavras em *-es.*

Grupo consonantal é a reunião de duas ou mais consoantes no interior de uma palavra. Os grupos consonantais podem ser classificados como:

(a) homogêneos quando formados por duas consoantes iguais ou geminadas: *stuppa, bucca, flamma, ille* etc.;
(b) heterogêneos quando formados por consoantes diferentes: *persona, clave, lacte, flumen, tribu* etc.;
(c) latinos quando já existiam na língua latina: *stuppa, clave, tribu* etc.
(d) românicos quando resultaram da síncope de uma vogal: *laborare > labrar > lavrar; apicula > apicla > abelha* etc.;
(e) próprios quando são formados de uma consoante oclusiva ou [f] mais uma líquida ([l] ou [r]): *placere, premere, flamma* etc.
(f) impróprios quando são formados por outras consoantes que não as descritas acima: *lt (multum); gn (regnu); pt (aptu)* etc.

Os grupos consonantais homogêneos se reduziram a consoantes simples:

bb > b: *sabbatu > sábado*
cc > c: *bucca > boca*
dd > d: *additione > adição*
ff > f: *effectu > efeito*
gg > g: *aggravare > agravar*
ll > l: *illa > ela*
mm > m: *flamma > chama*
nn > n: *pannu > pano*
pp > p: *stuppa > estopa*
tt > t: *gutta > gota*

Os grupos latinos mediais *-rr-* e *-ss-* se conservaram na ortografia, mas não correspondem a consoantes geminadas, que não existem em português: o RR e o SS são meros dígrafos, isto é, duas letras que representam um único som. Existem diversas possibilidades de realização sonora do que se escreve RR. Já o SS representa o fonema surdo [s] entre vogais.

O estudo dos grupos próprios depende de sua posição inicial ou medial.

Os grupos próprios iniciais formados de oclusiva ou [f] com a vibrante [ɾ] não se alteraram:

braciu > *braço*	*frenu* > *freio*
breve > *breve*	*gradu* > *grau*
cruce > *cruz*	*granu* > *grão*
credere > *crer*	*pratu* > *prado*
dracone > *dragão*	*probare* > *provar*
fructu > *fruto*	*truncu* > *tronco*

Os grupos próprios formados com a lateral [l] passaram pelas seguintes transformações:

(a) os grupos *cl*, *fl*, *pl* se palatizaram em *ch*:

clave > *chave*	*flagrare* > *cheirar*
clamare > *chamar*	*pluvia* > *chuva*
flamma > *chama*	*plenu* > *cheio*

(b) Essa transformação, evidentemente, compreendeu estágios intermediários. No caso de *cl-* > *ch-*, por exemplo, é provável que tenha havido a seguinte cadeia de mudanças: a primeira etapa desse processo de palatização — kl- > kɪ- — se verifica no italiano, em que o latim *clave* resultou em *chiave* ['kɪave]. Conforme nos explica Viaro (2011: 177):

> Em espanhol, esses encontros, no início dos vocábulos, promoveram a aférese da primeira consoante: lat *pluviam* > esp *lluvia*, lat *clavem* > esp *llave*, lat *flammam* > esp *llama*. No Noroeste peninsular [Galiza], a mudança foi mais complexa, uma vez que o *[ʎ] se ensurdeceu por assimilação parcial da consoante inicial, transformando-se, provavelmente, num som alveolar lateral fricativo e surdo *[ɬ] [...] Por outro lado, *p-*, *c-* e *f-* perderam totalmente seus traços articulatórios originais, adquirindo, por assimilação, os traços alveolares da lateral, convergindo, assim, numa única africada (semelhante ao que existe na língua nauatle, do México): *[pɬ] > *[tɬ], *[kɬ] > *[tɬ], *[ɸɬ] > *[tɬ]. Esse som instável transformou-se em [ʧ], grafado *ch* e assim permaneceu com grande vigor, até a época do Renascimento, quando port ant [ʧ] > port mod [ʃ].

Em galego, o que se escreve *ch* até hoje se pronuncia como [ʧ].

(c) Esses mesmos encontros passaram por outra transformação, a rotacização, em que a lateral [l] se torna a vibrante [ɾ]: *plaga* > *praga*; *flaccu* > *fraco*; *clavicula* > *cravelha*; *placere* > *prazer*. Essas formas com [ɾ] surgiram em fase posterior da língua e em camadas sociais diferentes daquelas em que se deu a transformação desses grupos em *ch*. Note-se que esse mesmo fenômeno ocorre fartamente no PB, em diversas variedades regionais e sociais sendo, no entanto, altamente estigmatizado pelos falantes urbanos letrados.

(d) os grupos *bl* e *gl* se transformaram em *br* e *gr*, ou se reduziram a *l*:

blandu > *brando*	*glattire* > *latir*
glute > *grude*	*blastemare* > *lastimar*

Os grupos próprios mediais formados de CONS + [ɾ], quando precedidos de outra consoante, não se alteraram: *membru* > *membro*; *mostrare* > *mostrar*; *scribere* > *escrever*; *exfricare* > *esfregar*.

(a) Quando precedidos de vogal, a primeira consoante é tratada como intervocálica e, assim, pode sonorizar-se, vocalizar-se, sofrer síncope ou conservar-se, de acordo com os processos de mudança que afetaram as consoantes em geral. Isso se deve ao caráter da vibrante [ɾ] que, como soante, compartilha alguns traços das vogais, e do [g] que se abranda comumente na semivogal [ɪ]:

latrone > *ladrão* → sonorização de surda intervocálica
lacrima > *lágrima* → sonorização de surda intervocálica
integru > *inteiro* → vocalização de [g]
regnu > *reino* → vocalização de [g]
libru > *livro* → degeneração de [b]
quadraginta > *quaraenta* > *quarenta* → síncope de sonora intervocálica

(b) Em alguns casos, o *r* da sílaba átona desapareceu por dissimilação total:

aratru > *arado* *rostru* > *rosto*
fratre > *frade* *cribru* > *crivo*

(c) No caso das palavras *padre* (< *pater*) e *madre* (< *mater*), que deram *pai* e *mãe*, há duas hipóteses. Conforme a primeira, as palavras *padre* e *madre* deram *pai* e *mãe* através das formas *pade* e *made*, de caráter afetivo (ou infantil). Compare-se com a forma afetiva popular brasileira *cumpade* e *cumade*, por *compadre* e *comadre*. A segunda hipótese é a de que já existiam no latim vulgar as formas *pate* e *made*, analógicas de *frate* ('irmão'), em que o *r* caiu por dissimilação total. A nasalidade de *mãe* se deve à consoante inicial.

Quanto aos grupos próprios mediais formados de CONS + [l]:

(a) *cl*, *fl*, *pl* palatizaram-se em *ch* quando precedidos de consoante, e em *lh* quando precedidos de vogal:

masculu > *masclu* > *macho*
fasculu > *fasclu* > *facho*
inflare > *inchar*
implere > *encher*
macula > *mancula* > *mancla* > *mancha*

scopolu > *iscoplu* > *escolho*
oculu > *oclu* > *olho*
auricula > *ouricla* > *orelha*
apicula > *apicla* > *abelha*
macula > *macla* > *malha*

(b) *bl*, *gl*, *tl* também se palatizaram em *lh*, quando antecedidos de vogal:

tribulu > *triblu* > *trilho*
tegula > *tegla* > *telha*
coagulare > *coaglar* > *coalhar*
vetulu > *vetlu* > *velho*
rotula > *rotla* > *rolha*

(c) Os mesmos estratos sociais que operaram a transformação desses grupos com [l] em grupos com [ɾ] quando iniciais também são responsáveis pela mesma trans-

formação desses grupos quando mediais: *nobile* > *noble* > *nobre*; *regula* > *regla* > *regra*; *obligare* > *obrigar*; *diabolu* > *diablo* > *diabro* (do arcaico *diabro* derivam as formas *diabrete*, *diabrura*, *endiabrar*).

(d) O grupo medial *tl* provavelmente já tinha se transformado em *cl* no latim vulgar, uma vez que o *Appendix Probi*, lista de prescrições sobre como pronunciar corretamente o latim, corrige: "Vetulus non veclus". A partir de *cl* deve ter ocorrido a sequência *cl* > *gl* > *il* > *li* > *lh*:

veclu > **veglo* > **veilo* > **velio* > *velho*.

(e) O mesmo vale para as palavras que em latim apresentavam *g* seguido de *n*, *r*, *l*, ou seja, soantes, aparentadas às vogais. Nessas condições, o *g* se vocalizou em [i] e a sequência descrita acima se desdobrou aqui também:

regnu > *reino*
integru > *inteiro*
tegula > **tegla* > **teila* > **telia* > *telha*
ligna > **leina* > **lenia* > *lenha*

Com relação aos grupos impróprios, ocorreram as seguintes mudanças:

(a) Os grupos consonantais iniciados por *s-* impuro, isto é, sem apoio de vogal inicial, recebem um *e* protético, sendo que no grupo *sc* seguido de vogal *e* ou *i*, o *s* sofre aférese:

scutu > *escudo*
scorpione > *escorpião*
sponsa > *esposa*
spatio > *espaço*
stella > *estrela*
stare > *estar*
scientia > *ciência*
scena > *cena*

(b) Nos grupos de oclusivas com outras consoantes, temos:

ps assimila-se: *ipse* > *esse*; *gypsu* > *gesso*

kt vocaliza-se: *nocte* > *noite*; *octu* > *oito*; *factu* > *feito*

pt vocaliza-se: *conceptu* > *conceito*; *acceptu* > aceito
ou assimila-se: *septe* > *sette* > *sete*; *captare* > *cattare* > *catar*

ks (=x) passa a *ix*, *is* ou *ss*: *saxu* > *seixo*; *mataxa* > *madaixa* > *madeixa*; *laxare* > *leixar* (arc.); *sex* > *seis*; *dixi* > *disse*

(c) Nos grupos resultantes dos prefixos *ad-*, *sub-*, *ab-* ocorre com muita frequência a assimilação: *substare* > *sustar*; *subjectu* > *sujeito*; *subterrare* > *soterrar*; *adversu* > *avesso*. Raramente se dá a vocalização: *absentia* > *ausência*.

(d) Nos grupos de constritivas mais outra consoante, temos:
rs e *mn* sofrem assimilação: *persicu* > *pêssego*; *persona* > *pessoa*; *somnu* > *sono*; *autumnu* > *outono*

rb e *lb* → o *b* se degenera em *v*:	*arbore > árvore; turbare > turvar; carbone > carvão; albu > alvo*
ns → síncope do *n*:	*mensa > mesa; ansa > asa; pensu > peso; mense > mês*
sc, seguido de *e* ou *i*, assimila-se:	*patescere > padecer; merescere > merecer*
ou passa a *ix*:	*pisce > peixe*
lt, *lc*, *lp* → vocalização do *l*:	*alteru > altru > outro; multu > muito; falce > fauce > foz; palpo > poupo*

(e) Os grupos *ml* e *mr*, resultantes da síncope de uma vogal, desenvolvem uma consoante de transição *b* (bilabial como *m*):

*simulante > *sem'lante > semblante*
*memorare > *mem'rar > membrar > lembrar*
*umeru > *om'ru > ombro*

Grupos de consoante mais semivogal. A semivogal [ɪ] chamada *iode* (de seu nome em hebraico, *yod*), exerceu grande influência na fonética histórica do galego e do português, gerando fonemas palatais e sibilantes desconhecidos na língua latina:

(a) Na palavra latina *miliu* existiam três sílabas: *mi-li-u*. Esse hiato final será desfeito, transformado em ditongo, no qual o *i* passa a ser uma semivogal: *mi-liu*. Em fase posterior, o grupo *li* (+ vogal) se transformará na consoante palatal escrita *lh* [ʎ]. O mesmo ocorreu com o grupo *ni* (+ vogal), que se transformará na consoante palatal escrita *nh* [ɲ]. As grafias *lh* e *nh*, em português, foram tomadas de empréstimo ao provençal. Assim, temos:

fi-li-u > fi-lio > filho
pa-le-a > pa-lia > palha
con-si-li-u > con-si-lio > conselho
ci-co-ni-a > ci-co-nia > cegonha
ve-re-cun-d-ia > ve-re-cun-nia > vergonha
te-ne-o > te-nio > tenho
li-ne-a > li-nia > linha
se-ni-o-re > se-nio-re > senhor
a-ra-ne-a > a-ra-nia > aranha

(b) Os grupos *ci* e *ti* se tornam as sibilantes [ts] ou [dz] grafadas *ce*, *ci*, *ç* e *z*, respectivamente. Mais tarde, em português, seriam pronunciadas sem a oclusão: [s] e [z]:

facio > *faço* ['fatso]
lancea > *lancia* > *lança* ['lãtsa]
iudiciu > *judicio* > *juízo* [ʒu'idzo]
iaceo > *jacio* > *jazo* ['ʒadzo]
pretiu > *preço* ['pretso] ~ *prezo* ['prɛdzo]
gratia > *graça* ['gratsa]
iustitia > *justiça* [ʒus'titsa] ~ *justeza* [ʒustedza]
palatiu > *paaço* > *paço* ['patso]
capitia > *cabitia* > *cabeça* [ka'betsa]

(c) No galego, como já mencionamos, [ts] e [dz] se transformaram na interdental [ɵ] como também ocorreu em castelhano.

(d) O grupo *di-* se assimila em [ts] (mais tarde [s]) ou se palatiza em [ʒ]:

audio > ouço ['outso]
ardeo > ardio > arço (arc.) ['artso]
verecundia > verecundya > vergonça (arc.) [ver'gõtsa]
insidio > ensedio > ensejo
hodie > hoje
invidia > inveja

(e) O grupo *gi-* se palatiza em [ʒ]:

fugio > fujo
*angelu > *angeo > *angiu > anjo*
spongia > esponja

(f) Nas variedades de alguns grupos sociais, as consoantes *d* e *g* dos grupos *di-* e *gi-* foram tratadas como intervocálicas, sofrendo, portanto, síncope como as sonoras intervocálicas em geral:

radiu > radio > raio
badiu > badio > baio
navigiu > navigio > navio
exagiu > exagio > ensaio

(e) Os grupos *si* e *ssi* passam, respectivamente, a *ij* e *ix*, depois de uma metátese e de uma palatização da consoante por influxo do iode:

*baseu > *basiu > *bai[z]u > *baijo > beijo*
*caseu > *casiu > *cai[z]u *caijo > queijo*
ecclesia > igreija (arc.) > igreja
*russeu > *russiu > *ruixu > roixo (arc.) > roxo*
*passione > *paisson > *paixon > paixão*
**bassiare > *bassiar > *baissar > baixar*

(f) No português brasileiro atual, os grupos escritos *eij*, *eix*, *aix* tendem a se simplificar, por assimilação do iode, em *ej*, *ex*, *ax*, como em *beijo* ['beʒu], *queixo* ['keʃu], *caixa* ['kaʃa] etc. O mesmo vale para o grupo *eir*: *cheiro* ['ʃeru]. Na variedade lisboeta do português europeu, por efeito de fenômeno oposto, o da dissimilação, a vogal [e] do grupo escrito *ei* se transformou numa vogal média central aberta [ɐ], donde as pronúncias *beijo* ['bɐɪʒu], *cheiro* ['ʃɐɪru] etc., com preservação do iode. Nos demais dialetos do português europeu, também ocorreu a mesma assimilação que se verifica no PB. No entanto, como a variedade lisboeta é tomada como norma, os falantes das demais variedades tendem a adotar a pronúncia [ɐɪ] do ditongo.

7.6 Do português europeu ao português brasileiro

Trazido para terras brasileiras a partir do século XVI pelos navegadores e colonos lusitanos, o português enfrentou as vicissitudes históricas que estudamos no capítulo 5,

entrando em concorrência e em contato com as línguas indígenas e, mais tarde, com as línguas africanas. No período da ocupação do território pelos portugueses, a língua se encontrava num período de transição entre sua fase medieval e sua fase clássica.

A fase clássica tem esse nome porque representa um momento de grande *investimento* cultural e político na língua por parte dos intelectuais portugueses da época. Acompanhando a expansão marítima e a constituição de uma potência imperial, os letrados portugueses se esforçaram por conferir à língua elementos morfossintáticos e lexicais que permitissem a ela desempenhar todas as funções esperadas de uma língua oficial, adequada à elaboração das leis, à produção de conhecimento científico, à institucionalização educacional.

Nessa fase, o gênio poético de Camões se ergue acima de todos os demais escritores portugueses e sua produção literária vai marcar definitivamente os destinos da modalidade literária da língua. Sua obra de propaganda ideológica da vocação imperialista de Portugal, *Os Lusíadas* (1572), é o marco simbólico que inaugura a fase clássica da língua. O poema se inspira diretamente na epopeia de Virgílio, *A Eneida* (século I a.C.), produzido também como propaganda ideológica do poderio imperial romano em sua fase de maior desenvolvimento — não por acaso, Virgílio é considerado o maior nome da fase *clássica* do latim. O vínculo direto com Virgílio é assumido no primeiro verso da epopeia camoniana:

Arma virumque cano ("Eu canto as armas e os varões")	*As armas e os barões assinalados* [...] *Cantando espalharei por toda parte*

Também nessa fase, acompanhando um movimento intelectual difundido por toda a Europa ocidental, são produzidas as primeiras gramáticas da língua portuguesa, igualmente marcadas por um compromisso ideológico com a expansão imperial. Defendendo o uso do português, e não mais do latim, como língua nacional, oficial e colonizadora, o primeiro gramático, Fernão de Oliveira, escreve explicitamente: "*Melhor é que ensinemos a Guiné do que sejamos ensinados de Roma*" (1536). Essas palavras ecoam as de Antonio de Nebrija, autor da primeira gramática do castelhano: "*Siempre fue la lengua compañera del imperio*" (1492) [▶143]. As primeiras obras de investigação e prescrição do português foram:

1536: *Gramática da linguagem portuguesa*, Fernão de Oliveira;
1540: *Gramática da língua portuguesa*, João de Barros;
1574: *Regras que ensinam a maneira de escrever*, Pero de Magalhães de Gândavo;
1576: *Ortografia da língua portuguesa*, Duarte Nunes de Leão;
1606: *Origem da língua portuguesa*, Duarte Nunes de Leão.

No entanto, assim como as províncias romanas não foram colonizadas pelos intelectuais que usavam o *latim clássico*, literário e erudito, mas por soldados, agricultores e funcionários que falavam o *latim vulgar*, também no caso da expansão colonial portu-

guesa não foram os pouquíssimos conhecedores da leitura e da escrita e da norma literária recém-estabelecida os que vieram ocupar as terras conquistadas. Marinheiros, colonos, degredados, criminosos comuns, perseguidos religiosos, pequenos funcionários etc. eram falantes de diferentes variedades populares do português europeu da época.

Uma vez aqui implantada, ainda que precariamente no início da colonização, a língua começou a seguir destinos diferentes dos que seriam seguidos pelos falantes da Europa, dando início à formação do PB, língua que hoje apresenta uma *gramática* bastante distinta da do português europeu, o que nos autoriza a dizer que se trata de *uma língua diferente*, a mais falada do grupo portugalego e uma das mais faladas do mundo. Somente a ideologia nostálgica da colonização e de depreciação do que é propriamente brasileiro explica a subserviência da nossa tradição gramatical e filológica com relação ao português europeu e à negação do *status* de *língua* ao PB.

7.7 Vocalismo brasileiro

No que diz respeito às vogais tônicas, o PB não apresenta grande variação. Uma exceção notável, que vamos ver mais abaixo, é a desnasalização das tônicas na variedade paulistana. A sílaba tônica é a que guarda, por assim dizer, a identidade de uma palavra, daí a forte tendência a preservá-la dos efeitos da mudança linguística. Quando examinamos a história das línguas românicas, por exemplo, fica evidente que, por maior que seja a erosão fonética sofrida por uma palavra em seu processo de mudança, a sílaba tônica na maioria das circunstâncias se conserva:

ásinu > *asno* (cf. francês *âne*)
ánima > *alma* (cf. francês *âme*)
fábrica > *frágua*
hómine > *homem*
legítimu > **leídimu* > *lídimo* > *lindo*
límite > *linde*
régula > *regra* ~ *régua* ~ *relha*

Essa mesma tendência à preservação da tônica se verifica em processos de mudança ocorridos em diversas variedades regionais do PB:

árvore > *arvre* ~ *arve*
córrego > *corgo*
cubículo > *cuvico*
imundície > *mundiça*
lâmpada > *lampa*
pássaro > *passo*
sábado > *sabo*
víbora > *briba* ('lagartixa')

Na variedade carioca, as vogais tônicas passam atualmente por um processo de semi-ditongação, proveniente de uma pronúncia mais longa dessas vogais:

alô → [a'lo^{a}]
cinco → ['sĩ^{ã}ku]
disse → ['ʤi^{i}si]
fiz → ['fi^{i}ʃ]
lindo → ['lĩ^{ã}du]
nove → ['no^{a}vi]

Ocorre também uma ditongação na variedade paulistana depois da vogal tônica nasal [ẽ] seguida de travamento silábico:

fazenda → [fa'zẽ^{ẽ}da]
momento → [mo'mẽ^{ẽ}tu]
entendendo → [ĩtẽ^{ẽ}'dẽ^{ẽ}du]

Também na variedade paulistana, decerto devido ao contato de falantes de dialetos italianos com o PB, ocorre a não nasalização de vogais tônicas diante de [m] e [n], ao contrário do que se dá em todas as demais variedades da língua:

Antônio → [ã'tɔniʊ] (cf. RJ, MG, BA etc.: [ã'tõniʊ]
fome → ['fɔmi] (cf. RJ, MG, BA etc.: ['fõmi])
homem → ['ɔmẽ^{ẽ}] (cf. RJ, MG, BA etc. ['õmẽ^{ẽ}])
viemos → [vi'ɛmus] (cf. MG: [vi'ɛ̃mus])

Em muitas variedades do PB, ocorre uma ditongação das vogais tônicas finais (palavras oxítonas) quando seguidas de [s]. Esse fato é tão comum que a não ditongação é considerada como traço "regional" (por exemplo, mineiro ou gaúcho):

depôs → [de'pɔɪs] ~ [de'pɔiʃ]
fez → ['feɪs] ~ ['feiʃ]
mas → ['maɪs] ~ ['maiʃ]
nós → ['nɔɪs] ~ ['nɔiʃ]
paz → ['paɪs] ~ ['paiʃ]
rapaz → [ha'paɪs] ~ [ha'paiʃ]
reduz → [he'duɪs] ~ [he'duiʃ]

Com isso, pares de palavras como *mas/mais*, *pôs/pois*, *paz/pais* etc. só diferem na grafia, uma vez que têm a mesma pronúncia. Daí a frequente hesitação de muitas pessoas na hora de escrever *mas* ou *mais*.

Na variedade carioca, esse processo avançou um pouco mais e se verifica também nas vogais tônicas em sílabas travadas por [s] e [z] de palavras não oxítonas. É uma mudança que ocorreu entre os falantes das gerações mais novas contemporâneas (por volta dos 25 anos de idade):

asma → ['aɪʒma]
custo → ['kuɪʃtu]
festa → ['fɛɪʃta]
posto → ['poɪʃtu]

No que diz respeito às átonas finais, as três principais línguas do grupo portugalego apresentam atualmente sistemas diferentes:

(a) galego: [a] [e] [o] → conservação da pronúncia arcaica:
casa ['kaśa] *torre* ['toře] *oco* ['oko]

(b) português brasileiro: [a] [i] [u] → conservação da pronúncia medieval:
casa ['kaza] *torre* ['tohi] *oco* ['oku]

(c) português europeu: [ɐ] [ə] [u] → inovações surgidas no período moderno:
casa ['kazɐ] *torre* ['toʀə] *oco* ['oku]

Em algumas variedades do PB, concentradas na região Sul, ocorre a realização de *-e* final como [e] e não como [i] É famosa a frase estereotipada atribuída aos curitibanos: *leite quente dói no dente da gente*.

Sobre as vogais pretônicas, ocorre grande variação no PB:

(a) Nas variedades do Nordeste e do Norte, uma das principais características é a abertura das vogais pretônicas [e] e [o] que, nas demais variedades, só ocorrem em sílabas tônicas:
cooperar → [kɔpɛ'ɾa(h)]
problema → [prɔ'blẽma]
recepção → [hɛsepi'sɐ̃ʊ̃]
televisão → [tɛlɛvi'zɐ̃ʊ̃]

(b) Em muitas variedades, é comum o alçamento das pretônicas [e] e [o]. Um fator que favorece esse fechamento é a presença de uma consoante bilabial ([b], [m], [p]):
ameaça → [ami'asa]
besouro → [bi'zoɾu]
bolacha → [bu'laʃa]
boneca → [bu'nɛka]
bonito → [bu'nitu]
melhor → [mi'ʎɔh]
molambo → [mu'lãbu]
moleque → [mu'lɛki]
pequeno → [pi'kẽnu]
poeira → [pu'eɾa]

(c) Outro fator que favorece o fechamento é a harmonização vocálica, isto é, a assimilação da pretônica por outra vogal fechada na sílaba seguinte, principal-

mente tônica. Em muitos casos, essa harmonização é ainda mais favorecida pela presença de uma consoante bilabial:

harmonia → [afimu'nia]
medida → [mi'ʤida]
menino → [mi'nĩnu]
mosquito → [mus'kitu]
pedido → [pi'ʤidu]
perua → [pi'ɾuʊa]
fortuna → [fuh'tũna]
gordura → [gufi'duɾa]
dormir → [dufi'mih]
cozinha → [ku'zĩɲa]
devia → [ʤi'via]
Severino → [sivi'ɾĩnu]

(d) Ocorrem também outros fechamentos das pretônicas [e] e [o] difíceis de sistematizar e atribuíveis principalmente à origem geográfica do falante. Os fechamentos são mais comuns no Sudeste (com exeção de São Paulo), no Nordeste e no Norte, e menos frequentes em São Paulo e na região Sul:

colégio → [ku'lɛʒʲu] (RJ, MG, BA) ~ [ko'lɛʒʲu] (SP)
fogão → [fu'gɐ̃ʊ̃] (RJ, MG, BA) ~ [fo'gɐ̃ʊ̃] (SP, RS)
futebol → [fuʧi'bɔʊ] (RJ, MG, BA) ~ [fute'bɔʊ] (SP)
português → [puxtu'geiʃ] (RJ) ~ [pɔhtu'geiʃ] (PE) ~ [pɔɾtu'geɪs] (SP)
teatro → [ʧi'atɾu] (RJ, MG, BA) ~ [ti'atɾu] (PE, PB) ~ [te'atɾu] (SP, RS)

(e) Na variedade mineira da Zona da Mata e região de Belo Horizonte, ocorre uma pretônica [ɐ] isto é, um "a" centralizado e fechado em sílaba pretônica inicial de palavras longas:

abacaxi → [ɐbaka'ʃi] ~ [ɐbɐkɐ'ʃi]
amadurece → [ɐmadu'ɾɛsi] ~ [ɐmɐdu'ɾɛsi]
aparecida → [ɐpari'sida] ~ [ɐpɐɾi'sida]
maravilhosa → [mɐɾavi'ʎɔza] ~ [mɐɾɐvi'ʎɔza]

As vogais nasais ocorrem, no PB mais geral, em sílaba (tônica ou átona) travada por consoante nasal, em ditongos nasais (tônicos ou átonos) e em vogal tônica sucedida por uma consoante nasal:

acampar → [a'kɐ̃pah] (sílaba átona travada)
casaram → [ka'zaɾɐ̃ʊ̃] (ditongo nasal átono)
casarão → [kaza'ɾɐ̃ʊ̃] (ditongo nasal tônico)
cano → ['kɐ̃nu] (vogal tônica seguida de consoante nasal)

Em variedades nordestinas, a nasalidade decorrente da presença de uma consoante nasal depois da vogal ocorre também em sílabas átonas, ao contrário do que se dá nas demais variedades:

banana → [bɐ̃'nɐ̃na] (NE) ~ [ba'nɐ̃na] (SE)
camarada → [kɐ̃ma'ɾada] (NE) ~ [kama'ɾada] (SE)
amei → [ɐ̃'meɪ] (NE) ~ [a'meɪ] (SE)

No PB ocorrem os seguintes ditongos:

(a) ditongos orais decrescentes:

[aʊ] → *mau, mal*
[eʊ] → *meu, feltro*
[ɛʊ] → *céu, mel*
[iʊ] → *viu, mil*
[oʊ] → *gol, golfo*
[ɔʊ] → *sol, golpe*
[uʊ] → *sul, multa*

[aɪ] → *pai, raiva*
[eɪ] → *sei, peito*
[ɛɪ] → *anéis*
[iɪ] → *fiz, visto* (RJ)
[oɪ] → *foi, foice*
[ɔɪ] → *dói*
[uɪ] → *fui, circuito*

(b) ditongos orais crescentes:

[ɪa] → *glória*
[ɪi] → *série, cárie*
[ɪo] → *funcionar*
[ɪu] → *mistério*

(c) ditongos nasais:

[ãʊ̃] → *pão, falam*
[ãɪ̃] → *mãe, cãibra*
[ẽɪ̃] → *alguém, falem; vento* (SP)
[ĩĩ] → *sim* (RJ)
[õɪ̃] → *põe*
[ũɪ̃] → *muito, ruim*

Acerca dos ditongos no PB, algumas observações merecem ser feitas:

(a) Os ditongos escritos *ai* e *ei* se transformam nas vogais simples [a] e [e] quando precedem uma consoante palatal. Ocorre, portanto, a assimilação do [ɪ], palatal, com a consoante:

beijo ['beʒu]
caixa ['kaʃa]
faixa ['faʃa]
peixe ['peʃi]
queijo ['keʒu]

(b) O ditongo *ei* também se simplifica diante de [ɾ]:
brasileiro [brazi'leɾu]
cabeleireiro [kabile'ɾeɾu]
cheiro ['ʃeɾu]
feira ['feɾa]
inteiro [ĩ'teɾu]

Essa monotongação generalizada leva muitas pessoas a querer escrever, por hipercorreção, as palavras *bandeja, prazeroso, caranguejo* etc. como "bandeija", "prazeiroso", "carangueijo" etc.

(c) Por outro lado, em variedades do Nordeste, o ditongo [eɪ] pode ocorrer sem sua semivogal. Isso se deve, decerto, ao caráter mais conservador dessas variedades, que preservaram as formas com hiato, anteriores ao surgimento da semivogal:
areia [a'ɾea] (NE) ~ [a'ɾeɪa] (SE)
feia ['fea] (NE) ~ ['feɪa] (SE)
ideia [i'dɛa] (NE) ~ [i'dɛɪa] (SE)

(d) Por sua vez, o ditongo escrito *ou* é sempre pronunciado [o], em todas as variedades e independente do contexto fonético em que se encontre:
ouro ['oɾu] → inicial tônica
tesouro [ti'zoɾu] → medial tônica
acabou [aka'bo] → final tônica
couraça [ko'ɾasa] → inicial átona
agourar [ago'ɾah] → medial átona

A pronúncia do ditongo só ocorre em situações de fala monitorada, sobretudo em ocasiões de oralização de texto escrito, em contexto formal.

(e) A monotongação generalizada de *ou* conduziu as formas verbais com [o], tônico fechado, para [ɔ], tônico aberto, por analogia com os pares em que o verbo tem vogal tônica aberta e o substantivo, vogal tônica fechada:

SUBST. [o]	VERBO [ɔ]		SUBST. [o]	VERBO [ɔ]
acordo	*acordo*	*roubo*	['hobu]	['hɔbu]
almoço	*almoço*	*estouro*	[ist'oɾu]	[is'tɔɾu]
jogo	*jogo*	*pouso*	['pozu]	[pɔzu]

(f) A vocalização generalizada do [l] final em [ʊ] fez ressurgir, porém, o ditongo [oʊ], e seu correspondente aberto [ɔʊ], que só ocorrem em palavras escritas com *ol*: *gol, golpe, sol, solto, volto, envolto, lençol* etc.

(g) O ditongo [uɪ] é frequentemente pronunciado como um hiato *u-i*, com sílaba tônica no *i* (embora a TGP condene essa pronúncia hiatizada):
circuito [sih'kuɪtu] ~ [sihku'itu]
fluido ['fluɪdu] ~ [flu'idu]
gratuito [gɾa'tuɪtu] ~ [gɾatu'itu]

(h) O ditongo nasal [ũĩ] só ocorre nas palavras *muito / muita / muitos / muitas*, por influência da nasal inicial, e no adjetivo *ruim* e seu derivado *ruindade*. A pronúncia de *ruim* como hiato [huĩ], apesar de prescrita pelos puristas, é cada vez mais rara no vernáculo do PB.

As descrições tradicionais da língua postulam a noção de "tritongo", com os exemplos clássicos de *Paraguai* e *saguão*. No entanto, preferimos adotar aqui a análise proposta por Cristófaro Silva (1999: 100), que trata de consoantes complexas. Como bem demonstra a autora, os chamados "tritongos" só ocorrem com as sequências *gu-* e *qu-*. Ora, o que está em jogo aqui não são as vogais, mas as consoantes, que

nesse caso não são [g] e [k] mas sim suas formas velarizadas [gʷ] e [kʷ]. Os latinos reconheceram o caráter excepcional da consoante [kʷ] ao atribuir a ela um símbolo gráfico especial, o Q, que só era usado seguido de *u*, empregado para indicar a velarização. Esse grupo escrito *qu* só ocorria diante de vogal. Ao contrário do que pode parecer, então, a letra *u* nesse caso não representa a semivogal [ʊ], mas o traço velarizador da consoante, o mesmo se aplicando à grafia *gu*. As consoantes [gʷ] e [kʷ] só ocorrem seguidas de vogal ou de ditongo.

(a) No PB as consoantes [gʷ] e [kʷ] podem ser pronunciadas sem a velarização em algumas palavras:
líquido → ['likʷidu] ~ ['likidu]
quatorze → [kʷa'tohzi] ~ [ka'tohzi]
questionar → [kʷesʧo'nah] ~ [kesʧo'nah]
quota → ['kʷɔta] ~ ['kɔta]
sanguíneo → [sã'gʷiniʊ] ~ [sã'giniʊ]

(b) Em muitas outras palavras, no entanto, o traço velar é conservado, sem variação:
quadro ['kʷadɾu]
quarto ['kʷahtu]
língua ['lĩgʷa]
água ['agʷa]
sequela [sekʷɛla]
tranquilo [trã'kʷilu]
aguentar [agʷẽ'tah]
sagui [sa'gʷi]
saguão [sa'gʷ͡ãʊ̃]
paraguaio [para'gʷaiu]
uruguaio [uɾu'gʷaiu]
quinquênio [kʷĩ'kʷẽniu]

Os *hiatos o-a* e *u-a* são evitados, em muitas variedades do PB, pela inserção de uma semivogal [ʊ]. Essa pronúncia é menos frequente nas variedades do Norte e Nordeste, mais conservadoras em geral no plano fonético. No Sudeste, porém, é muito comum:
boa ['boʊa] ~ ['boa]
pessoa [pe'soʊa] ~ [pe'soa]
rua ['huʊa] ~ ['hua]

Outro traço conservador característico das variedades nordestinas é a manutenção do hiato em palavras que o perderam devido ao fenômeno do sândi descrito acima [▶300]:
tenha ['tẽa] (NE) ~ ['tẽɲa] (SE etc.)
venha ['vẽa] (NE) ~ ['vẽɲa] (SE etc.)
vinho ['vĩu] (NE) ~ ['vĩɲu] (SE etc.)

7.8 Consonantismo brasileiro

No sistema consonantal do PB, alguns fatos merecem destaque, por caracterizarem a nossa língua em contraste com as demais do grupo portugalego.

De todos os sons da língua, aquele representado pela letra R (ou RR) é decerto o que apresenta maior variação, uma variação eminentemente geográfica e não social:

[ɾ] → é o *r* presente em *puro* e em *preto*, ou seja, entre vogais e depois de consoante; não apresenta variação regional, a não ser uma leve retroflexão em alguns falares "caipiras" (interior de SP sobretudo).

[r̆] → essa vibrante alveolar vozeada está presente em *rio* e *terra*, quando pronunciada com múltiplas vibrações da língua junto aos alvéolos. Rara no PB, a não ser em algumas variedades do Sul e em falantes de faixa etária avançada. É o *r* do castelhano e do italiano padronizados, como em *tierra* e *terra*.

[X] → pronúncia característica, por exemplo, da variedade carioca em *rio* e *terra*. Ocorre fricção audível na região velar. Se assemelha ao que se escreve com *j* em castelhano (*Juan, caja, jamón*).

[ɣ] → é a correspondente vozeada da consoante anterior e ocorre em final de sílaba seguida de consoante vozeada como em *perda*.

[ɹ] → é o conhecido "R caipira", do qual já falamos em [▶233]. Antigamente restrito às variedades rurais, já se expandiu para as zonas urbanas e hoje já faz parte da variedade paulistana, principalmente entre os jovens das periferias e da região metropolitana.

---

[h] → pronúncia característica, por exemplo, da variedade mineira de Belo Horizonte em *rio* e *terra*. Não ocorre fricção audível na região velar.

[ɦ] → correspondente vozeada da anterior, ocorre em final de sílaba seguida de consoante vozeada, como em *perda*.

Convém lembrar que em zonas de marcada imigração alemã é muito comum a não distinção entre o que se escreve R e RR, uma vez que nos dialetos alemães trazidos para cá essa distinção não ocorre. Com isso, palavras como *caro* e *carro* soam idênticas: ['karu]. Também nessas áreas (PR, SC e RS), é comum não haver distinção entre *queijo* e *queixo*, uma vez que os dialetos alemães desconhecem a palatal [ʒ], de modo que as duas palavras são pronunciadas ['keʃu]. Esses falantes se sentem estigmatizados pelos demais de sua região, principalmente das zonas metropolitanas, que produzem as distinções *caro/carro*, *queijo/queixo*.

A letra *l* também representa, no PB, pelo menos três sons bem distintos:

[l] → é o som presente duas vezes em *lula*. Como travamento silábico e/ou final de palavra, esse som lateral é cada vez mais raro no PB, ficando restrito a algumas variedades do Sul do país e a falantes mais idosos.

[ɫ] → representa uma velarização de [l]. Na grande maioria das variedades do PB, a consoante [l] em final de sílaba (e de palavra) se vocalizou em [ʊ]. No entanto, em algumas variedades, sobretudo no Sul e em faixas etárias mais avançadas, a vocalização ainda não se concluiu, permanecendo no estágio velarizado [ɫ]. Essa consoante é a mesma que ocorre em inglês e em português europeu nas mesmas condições de final de sílaba (e de palavra). Para nossos ouvidos, ela se assemelha a um "u" seguido de um levantamento da ponta da língua até os alvéolos.

[ʊ] → é a vocalização total do antigo [l], presente na ampla maioria das variedades do PB. Esse fenômeno fez ressurgir o ditongo [oʊ] e seu correspondente aberto [ɔʊ], que só ocorrem nas palavras escritas com *ol*: *gol, golpe, sol, solto*. Essa vocalização é responsável pela hesitação de muitos falantes, na hora de escrever, entre *l* e *u* (*mal/mau*), além de criar dificuldades para as crianças em processo de alfabetização. Esse fenômeno ocorreu no francês antigo, como revela a ortografia: *altu-* > *haut*; *saltu-* > *saut*; *malva* > *mauve*; *soldu-* > *sou*; *caldu-* > *chaud* etc.

A consoante [ɲ], escrita *nh*, tem pronúncia mais próxima de [nʲ] em algumas variedades, principalmente no Maranhão, no Pará e no Amapá. Nessas variedades, o verbo *sonha* tem pronúncia idêntica ao do nome próprio *Sônia*, palavras que nas outras variedades apresentam pronúncias distintas. Por outro lado, esses mesmos falantes palatalizam o [n] quando seguido de [i] em palavras como *bonito* e *aniversário* que soam como [bo'nʲitu] e [ãnʲivɛh'sariu], respectivamente.

A consoante [ʎ], escrita em geral *lh*, sofre deslateralização em incontáveis variedades do PB, reduzindo-se então à semivogal palatal [ɪ]:

folha ['foɪa]
palha ['paɪa]
telha ['teɪa]
trabalho [tra'baɪu]
velho ['vɛɪu]

Essa deslateralização é consequência natural das próprias características articulatórias da consoante [ʎ]. Conforme explica Lausberg (1981: 71): "Por afrouxamento e, finalmente, abandono da oclusão central forma-se do [ʎ] (difícil de pronunciar por causa da elasticidade reduzida do dorso da língua) muito naturalmente a fricativa palatal [ɪ]". Essa é a pronúncia padrão, por exemplo, do francês moderno, língua em que o [ʎ] desapareceu há alguns séculos. Também ocorre amplamente em diversas línguas da família espanhola (o peruano, o mexicano e o andaluz, por exemplo). Em galego, a pronúncia [ʎ] há muito desapareceu. O fenômeno é tão natural que muitos

estrangeiros, ao terem de aprender PB, realizam como [ɪ] a consoante [ʎ]. No Brasil, porém, esse é um dos traços fonéticos mais estigmatizados pelos falantes urbanos mais letrados, que tratam de evitá-lo em sua pronúncia, para não se identificarem com os falantes rurais ou rurbanos.

A consoante [ʎ] também pode vir escrita *-li-* ou, mais raramente, *-le-*. A grafia *-li-* é mera conservação da escrita etimológica latina (*família*, *mobília*, *Júlia*, *Brasília*). De fato, o nome do mês *julho* e o nome próprio *Júlio* têm a mesmíssima pronúncia. O substantivo *óleo* e o verbo *(eu) olho* têm também pronúncia idêntica.

Mais raramente, ao invés da deslateralização, ocorre a despalatalização da consoante [ʎ], como em *mulher*, pronunciado [mu'lɛ]. Com o pronome *lhe*, a despalatalização é generalizada a todo o VGB: [li].

Alguns falantes de variedades estigmatizadas inserem, por hipercorreção, um [ʎ] onde ele jamais existiu: *tocaia* [tu'kaʎa].

Na variedade de Belém do Pará (e provavelmente em outras do mesmo estado), a consoante [l] quando seguida de [i] sofre palatalização:

ali [a'ʎi]
galinha [ga'ʎĩn'a]
liga ['ʎiga]

Fenômeno devido ao contato do português brasileiro com as línguas africanas é a palatalização das consoantes [d] e [t] diante de [i]. Por ser amplamente distribuído pelo território brasileiro, e por constituir pronúncia de prestígio, essa palatalização não é repelida por ninguém. Ao contrário, a pronúncia [di] e [ti], sem palatalização, é que é sentida como "regional" (nordestina ou "caipira", por exemplo). Na região Nordeste, a variedade da cidade do Salvador se destaca por essa palatalização, ausente em Pernambuco, Alagoas, Paraíba e Rio Grande do Norte, além do sertão baiano, entre outros lugares. Decerto a forte presença demográfica do elemento africano na capital baiana explica esse fenômeno, o mesmo se podendo dizer da variedade carioca e também da mineira da antiga região do ouro, para onde foram levados grandes contingentes de escravos para a exploração do minério:

advogado [adʒivo'gadu] (RJ e MG)
dia ['dʒia] (SE) ~ ['dia] (NE)
entendi [ĩʧĩ'dʒi] (Belo Horizonte)
gente ['ʒẽʧi] (SE) ~ ['ʒẽti] (NE)
ritmo ['hiʧmu] (SP, RJ, MG, BA)
tia ['ʧia] ~ ['tia]

Esse fenômeno explica por que a palavra *estádio*, com o sentido de 'etapa de um processo', deixou de ser usada com esse sentido, substituída por *estágio*, com evidente palatalização da sílaba [di] e cruzamento semântico.

No chamado "dialeto cuiabano", as palatais [ʤ] e [ʧ] ocorrem como africatização de [ʒ] e de [ʃ], em todos os contextos. Esse fenômeno é atribuído pelos estudiosos ao substrato indígena bororo, povo que habitou originalmente a região de Cuiabá:

caixa ['kaʧa]
chapa ['ʧapa]
chuva ['ʧuva]
gente ['ʤẽte]
hoje ['oʤe]

A consoante escrita s pode ser palatalizada e pronunciada [ʃ] ou [ʒ] em travamento silábico e final de palavra, em muitas variedades do PB, com especial destaque para as regiões litorâneas. No caso do Rio de Janeiro, alguns pesquisadores atribuem a pronúncia à influência da fala de prestígio da corte portuguesa, quando se instalou na cidade em 1808, aumentando significativamente a população local. De fato, a palatalização do [s] em [ʃ], no português europeu, é documentada a partir do século XVIII, de modo que não ocorria na fala dos primeiros colonizadores. Em outras regiões (como Santa Catarina e Pará), é comum se atribuir a palatalização à imigração açoriana. Essa palatalização opõe, portanto, variedades muito próximas geograficamente, como a do Rio de Janeiro, de um lado, e a de Minas Gerais e São Paulo, do outro. É o estereótipo batizado de "s chiado", marca de identidade enfatizada por seus usuários, especialmente os cariocas:

estrela [ʃ'trela]
festa ['fɛʃta] ~ ['fɛɪʃta]
os mesmos [uʒ'meʒmuʃ] ~ [uɪʒ'meɪʒmuɪʃ]
os testes [uʃ'tɛʃʧiʃ] ~ [uɪʃ'tɛɪʃʧiʃ]

Em muitas variedades do PB, o índice pessoal a gente, pelo processo de erosão fonética típico da gramaticalização [▶170], é pronunciado [a'hẽt] ou [a'hẽʧ]. Por outro lado, em variedades nordestinas, a exclamação *ó gente!* se transformou em [ɔ'ʃẽti], assim como *virgem* se transformou em ['viʃi], com ensurdecimento da consoante [ʒ].

Em algumas variedades nordestinas, sobretudo no sertão, a consoante [v] se transforma na aspirada [h]:

cavalo [ka'halu]
tava ['taha]
vamo(s) ['hãmu] ~ ['hãm]
velho ['hɛiu] ~ ['hɛi]

Assim como no passado da língua o grupo *-rs-* sofreu assimilação e se tornou *-ss-* (*persona* > *pessoa*; *persicu* > *pêssego*; *adversu* > *avesso*), também no PB atual o mesmo fenômeno ocorre em muitas variedades:

cadarço [ka'dasu]
comércio [ko'mɛsiu]
universidade [univesi'dadʒi]

A pronúncia aspirada do R [h] decerto favoreceu essa assimilação: enfraquecida, a consoante, outrora vibrante múltipla, se tornou uma simples aspiração, que acaba por desaparecer diante da sibilante contínua [s]. O mesmo ocorre na palavra *murcho*, que muitos falantes pronunciam ['muʃu].

7.9 Metaplasmos brasileiros

Um dos postulados centrais da sociolinguística é o de que as mudanças ocorridas no passado continuam ocorrendo no presente. É o chamado princípio do uniformitarismo (ou *uniformitarianismo*), usado na geologia para designar o fato de que os processos geológicos do passado teriam ocorrido essencialmente com a mesma intensidade dos de hoje e, consequentemente, as modificações havidas na superfície terrestre teriam resultado de forças análogas às que agem no presente. Labov aplicou esse princípio à mudança linguística, com a intenção de mostrar que não há por que se surpreender com as transformações que as línguas sofrem atualmente, já que correspondem às mesmas causas e consequências das mudanças ocorridas no passado das línguas. Considerando a mudança linguística, conforme postulamos nessa gramática, como um fenômeno sociocognitivo, a ideia de que as mudanças se repetem de modo cíclico é perfeitamente lógica, uma vez que as capacidades cognitivas dos seres humanos têm sido as mesmas desde que a espécie se constituiu.

Vamos tratar aqui de alguns fenômenos de mudança que vêm ocorrendo no PB, com base nos metaplasmos que identificamos acima nas transformações do latim vulgar em galego e, em seguida, em português europeu.

A prótese é fenômeno constante na língua. Em diversas variedades do PB, por exemplo, é comum encontrarmos palavras acrescidas de um *a-* protético: *voar* > *avoar*; *lembrar* > *alembrar*; *sentar* > *assentar*; *repugnar* > *arripunar*; *pois* > *apois* etc. Muitos desses vocábulos são também formas arcaicas e clássicas conservadas em variedades regionais.

Ocorrem diversos casos de epêntese na língua atual. Para recuperar o padrão silábico CV (consoante + vogal), é comum a inserção de um [i] ou de um [e] depois da consoante chamada "muda": *pneu* > *p[i]neu* ~ *p[e]neu*; *football* > *futebol*; *advogado* > *a[dʒi]vogado* ~ *ad[i]vogado* ~ *ad[e]vogado*.

Em diversas variedades atuais do PB, ocorrem casos de suarabácti: *trem* > *terém*; *Clemente* > *Quelemente*; *Glória* > *Gulória*; *flor* > *fulô*. Alguns estudiosos veem nisso uma influência das línguas africanas trazidas para o Brasil. Para recuperar o padrão silábico CVCV, característico das línguas do grupo banto, os falantes recorreram ao suarabácti.

No aportuguesamento de vocábulos estrangeiros, é comum a paragoge, para evitar que a palavra termine em consoante: *club* > *clube*; *surf* > *surfe*; *chic* > *chique*. Em algumas variedades do PB, é comum a ocorrência de um [i] paragógico em palavras terminadas em [l] ou [ɾ]: *sol* > *sol*[i]; *cantar* > *cantar*[i]; *mar* > *mar*[i], fenômeno comum no português europeu e no galego popular.

A aférese ocorre com frequência em diversas variedades do PB atual, incluindo as urbanas de falantes mais letrados: *imaginar* > *maginar*; *aguentar* > *guentar*; *alcaguete* > *caguete*; *obrigado* > *brigado*; *estive* > *tive*.

Prosseguindo uma tendência muito antiga não só no galego e no português, mas também nas línguas românicas ocidentais, a síncope de sílabas postônicas de palavras proparoxítonas é regular em muitas variedades do PB: *árvore* > *arvre*; *pássaro* > *passo*; *fósforo* > *fosfre* ~ *frosfe*; *sábado* > *sabo*; *lâmpada* > *lampa*; *córrego* > *corgo*; *música* > ['muzga] etc.

Um caso de síncope frequente em variedades nordestinas, sobretudo sertanejas, é a eliminação de [ɾ] em encontros consonantais postônicos: *outro* ['otu]; *pobre* ['pɔbi]; *dentro* ['dẽtu].

Uma apócope muito difundida no PB é a supressão da consoante [ɾ] em final de palavra, sobretudo de infinitivos verbais: *cantar* > *cantá*; *vender* > *vendê*; *sair* > *saí*. Não causa rejeição da parte dos falantes urbanos letrados, que também falam assim. Em outras variedades, menos difundidas, é o [l] que sofre apócope: *qual* > *quá*.

Na língua atual ocorrem alguns casos de crase: *cooperar* > *coperar*; *álcool* > *alco*; *caatinga* > *catinga*; *feiíssimo* > *feíssimo* etc.

Por efeito da gramaticalização, algumas expressões sofreram sinalefa e se transformaram em palavras simples: *pelo amor de* > *pramode* ~ *prumode* ~ *mode*; *deixa eu ver* > *xovê*; *que é de* > *quedê* ~ *cadê* ~ *quede* etc.

Fenômenos de metátese e hipértese ocorrem com frequência na língua atual: *iogurte* > *iorgute*; *lagarto* > *largato*; *estupro* > *estrupo*; *tábua* > *tauba*; *dormir* > *dromir*; *vidro* > *vrido*; *prateleira* > *parteleira* etc.

O hiperbibasmo se manifesta no PB em pronúncias como as seguintes: *refém* > *réfem*; *clitóris* > *clítoris*; *rubrica* > *rúbrica*; *pudico* > *púdico*. Como as palavras proparoxítonas são quase todas de origem erudita, palavras paroxítonas têm sua acentuação alterada pela suposição de que é mais "correto" ou "sofisticado" o acento proparoxítono. É um caso típico de hipercorreção que, no entanto, pode provocar mudanças perenes na língua, como ocorreu com *pântano*, originalmente *pantânu*.

Já examinamos acima os casos da vocalização do [ʎ] em [ɪ] e do [l] em [ʊ].

Vimos também mais acima que o grupo inicial latino *ex-* se transformou em português em *enx-*. Essa tendência à nasalização prossegue nos dias atuais: muitos falantes pronunciam *exame* como "inzami", *exigente* como "inzigenti" e *exemplo* como "inzemplo" ou "inzempro", forma que existia, aliás, na fase arcaica da língua; "indiota" por *idiota*, "indentidade" por *identidade*. A produção do [i], vogal alta e fechada, próxima do palato, parece favorecer a nasalização. Isso explicaria a mudança do latim *sic* em *si* e posteriormente em *sim*. A mesma explicação talvez se aplique ao infinitivo *vir* que é pronunciado por praticamente todos os brasileiros como [vĩ], após a apócope do *-r* final dos infinitivos, também generalizada em todas as variedades do PB.

Por sua vez, a desnasalização é extremamente frequente, mesmo nas variedades urbanas de prestígio em contexto de interação menos monitorado, quando se trata de vogal nasal postônica:

cantaram [kã'taɾu]
fizeram [fi'zɛɾu]
homem ['õmi]
ontem ['õʧi]
viagem [vi'aʒi]

Um caso de abrandamento se verifica na pronúncia de *registro* como *re[z]istru*, que ocorre, por exemplo, na variedade soteropolitana, mesmo entre falantes letrados.

Já tratamos dos processos de palatalização das consoantes [d] e [t] no PB, assim como do [s]. Aqui vamos citar o caso de palatalização provocada pelo [ɪ] presente, não depois da consoante, mas antes dela, conforme ocorre em algumas variedades do Nordeste brasileiro:

muito ['mũʧu]
oito ['oʧu]
leite ['leʧi]

Observe-se que esse foi o mesmo fenômeno que ocorreu em castelhano, língua cujos falantes avançaram um passo no processo de palatalização do grupo latino *-ct-*:

multu > *muito* > **mutio* > *mucho* ['muʧo]
octu > *oito* > **otio* > *ocho* ['oʧo]
lacte > *laite* > *leite* > **letie* > *leche* [leʧe]

Processos de assimilação ocorrem frequentemente no PB atual. Nas variedades brasileiras em que se usa o pronome *tu*, ocorre a assimilação de *-st-* em *-ss-*: *viste* > *visse*; *fizeste* > *fizesse*; *foste* > *fosse*. Essa mesma assimilação explica a redução do par *este/esse* (e flexões) a *esse* (e flexões) [▶793]. Os ditongos ainda grafados *ou* e *ei* também se reduziram, por assimilação, a [o] e [e]: *pouco* > *p[o]co*; *roupa* > *r[o]pa*; *cheiro* > *ch[e]ro*; *beijo* > *b[e]jo* etc. Em algumas variedades específicas, também o ditongo *au* se reduz a [o]: *saudade* > *s[o]dade*; *Aurélio* > *[o]rélio*, como ocorreu fartamente no passado. Lembremos por fim a palavra *reiva* ['hɛiva] (< *raiva*), característica de algumas variedades rurais do PB.

Ocorre dissimilação (com eliminação) nas pronúncias atuais *próprio, registo* (culta no português europeu, assim como *rasto*). Em castelhano são da norma-padrão as formas *propio* e *orquesta*.

Já tratamos acima da metafonia associada à analogia no caso de pares como *o roubo / eu roubo*, em que se verifica a alternância [o] / [ɔ].

Fenômenos de rotacismo, que ocorreram na fase mais antiga da língua, são extremamente regulares na fala de milhões de brasileiros: *globo* > *grobo*; *placa* > *praca*; *inglês* > *ingrês*; *problema* > *pobrema* (com dissimilação eliminatória do primeiro *r*) etc. Existem variedades em que o *l* se torna *r* fora de encontros consonantais, mas em travamento silábico — é uma característica do chamado "dialeto caipira" (em que o *r* é retroflexo) e também de variedades sertanejas do Nordeste: *filme* > *firme*; *algum* > *argum*; *volta* > *vorta* etc.

Em algumas variedades, é o *r* que se torna *l*, fenômeno denominado lambdacismo. Esse *l* já se pronuncia como a semivogal [ʊ]: *terça* > *telça*; *carvão* > *calvão*; *garfo* > *galfo*; *serviço* > *selviço* etc.

Por fim, já registramos a não ocorrência do sândi interno em variedades nordestinas: *venha* ['vẽa].

7.1o A SÍLABA NO PORTUGUÊS BRASILEIRO

Os falantes do PB, como os de todas as línguas românicas, exibem ampla preferência pelo padrão silábico cv.cv (onde C = consoante, V = vogal). Na verdade, esse é o padrão silábico preferido pelos falantes da grande maioria das línguas humanas. Isso se deve, decerto, a fatores relacionados à articulação, um processo que exige esforço fisiológico: falar envolve inspirar e expirar, controlar o fôlego, ativar as pregas vocais, fazer a língua se mover para produzir os diferentes sons, abrir e fechar a boca, emitir sons pelo nariz etc. Não admira que depois de falar por horas seguidas uma pessoa se sinta cansada. Dois pesquisadores franceses nos explicam que

> concentrando-nos nos aspectos motores, a produção da fala requer energia, portanto a programação do controle motor tende a selecionar sequências de movimento que limitam as necessidades de energia. A primeira forma de minimizar o custo biomecânico do sistema articulatório é limitar deslocamentos de articuladores. (Rochet-Capellan / Schwartz, 2007: 1)

Por isso, é preferível iniciar uma palavra com algum tipo de obstrução — [b], [d], [g], [p], [t], [k], [m], [n], [f], [v] etc. — do que ter de fazer esse esforço no final da palavra. Daí uma tendência universal à eliminação de consoantes finais. Essa tendência é frequentemente acompanhada de outra: a não admissão de encontros consonantais numa mesma sílaba.

Alguns estudiosos até denominaram o padrão silábico CV como o "padrão ideal". Um grande número de línguas do mundo atingiram esse ideal, como por exemplo:

- o *japonês*: nessa língua não existem *encontros consonantais*; por isso, as palavras tomadas de empréstimo a outras línguas sofrem uma drástica adaptação ao padrão silábico japonês; por exemplo, o inglês *strike* ('greve') se transformou em *sutoraiku*; *Brasil* é *Burajiru*.
- as línguas da família *polinésia* (como o taitiano e o havaiano, entre muitas outras) só apresentam *sílabas abertas*, isto é, terminadas em vogal;
- o *malaio* e o *tagalo* (principal língua das Filipinas) seguem o padrão CV.CV;
- as línguas do grupo *tupi-guarani* também não admitem encontros consonantais (o português *cruz* se transformou em *curuçá* no tupi); a ortografia portuguesa pode criar a falsa impressão de sílabas travadas, como em *Pindamonhangaba*, que dividiríamos em *Pin.da.mo.nhan.ga.ba*; no entanto, as sílabas reais do tupi são *pi.nda.mo.nha.nga.ba*, porque nessa língua as consoantes [b], [d], [g] eram sempre precedidas de uma nasalização;
- o *iorubá* tem padrão CV e não tem encontros consonantais;
- as línguas do grupo *banto* também seguem o padrão CV e não conhecem os encontros consonantais;
- o *chinês* só admite algumas poucas consoantes no travamento silábico e não tem encontros consonantais;
- o *hauçá* (Sudão, Nigéria, Níger) também não apresenta encontros consonantais.

Pesquisas feitas com crianças em fase de aquisição da língua revelam que existe uma tendência à eliminação dos encontros consonantais e ao uso do padrão CV.CV, que reflete a mesma tendência encontrada na configuração das línguas humanas:

> Estudos comparativos da fonologia das línguas e das primeiras produções dos bebês humanos têm mostrado que *tendências na modelagem das línguas humanas* espelham *tendências no desenvolvimento da produção da fala*. Em primeiro lugar, CV é o tipo silábico predominante tanto no balbuciar canônico quanto nos inventários das línguas. Em segundo, as combinações CV com o mesmo lugar de articulação para a vogal e para a consoante são preferidas em ambos os casos. Em terceiro, as línguas preferem alternar o lugar de articulação para consoantes sucessivas em dissílabos CVCV [...] (Rochet-Capellan / Schwartz, 2007: 1).

Toda sílaba tem um núcleo, que é sempre uma vogal no PB. Ela também pode vir antecedida ou sucedida por uma semivogal (que vamos representar aqui pelo símbolo [v̯])— ou seja, o núcleo pode ser constituído por um ditongo crescente ou decrescente.

Esse núcleo pode ou não ser ladeado por duas ou mais consoantes. Como consoante inicial, podem ocorrer todas, com exceção de [ʎ], [ɲ] e [ɾ], que só ocorrem em posição medial. Palavras como *lhama*, *lhano* e *nhoque* são empréstimos, portanto, "corpos estranhos" no léxico do PB. O [ɾ], por sua vez, jamais ocorre em início de palavra (embora ocorra, por exemplo, no tupi e no italiano).

Se o núcleo for precedido por duas consoantes, elas têm de ser obrigatoriamente uma obstruinte (oclusiva ou fricativa) seguida de uma líquida ([l] e [ɾ]). É o que chamamos de encontro consonantal. No PB temos os seguintes encontros consonantais:

ENCONTRO CONSONANTAL	INÍCIO DE PALAVRA		MEIO DE PALAVRA	
	CCV	CCVV	CCV	CCVV
bl	blusa	blau	problema	
br	broto	breu	cobre	cobrei
cl	clima	claudicar	teclado	teclei
cr	crime	craibeira	micróbio	lacrei
dr	drama	druida	quadro	enquadrei
fl	fluxo	fluido	inflamar	inflei
fr	fruta	frei	resfriado	usufrui
gl	gládio	glauco	deglutir	
gr	grade	grei	regrado	imigrei
pl	pleno	plaina	aplicado	
pr	prata	preito	surpresa	comprei
tl	tláspio		atleta	
tr	truque	trauma	contrato	entrei
vr	vrido		lavrador	lavrei

[Obs.: o encontro [vl] só ocorre em alguns nomes próprios tomados de empréstimo a outras língua como Vladimir ou Vlado.]

Se no início da palavra podemos ter quase todas as consoantes (com as exceções já vistas: [ʎ], [ɲ] e [ɾ]), quando se trata de consoante pós-vocálica, as possibilidades se restringem

(1) às diversas realizações possíveis do que se escreve com a letra R ([ɾ], [h], [ɹ], [ɦ], [ř], [x], [ɣ]),
(2) com a letra S ([s], [ʃ], [ʒ]) e
(3) com a letra L ([l], [ɬ], [ʊ])[1].

Essas realizações vão diferir de acordo com a variedade linguística dos falantes e com o contexto fonético em que a consoante ocorre:

[1] Ao contrário de outros autores, considero que no PB existem, sim, vogais nasais. Por isso, não incluo entre as consoantes pós-vocálicas o chamado "arquifonema /N/". Dessa forma, considero as grafias *am, em, im, om, um, an, en, in, on, un* como dígrafos e não como uma sequência VC.

GRAFIA	EXEMPLOS		VARIEDADE LINGUÍSTICA	CONTEXTO FONÉTICO
R	CURTO	['kuɦtu] ['kuxtu] ['kuɾtu] ['kuɹtu]	mineira (BH), nordestina carioca paulistana, sulista "caipira"	antes de consoante desvozeada (surda)
	PERDA	['peɦda] ['peɣda] ['peɾda] ['peɹda]	mineira (BH), nordestina carioca paulistana, sulista "caipira"	antes de consoante vozeada (sonora)
S	CUSTO	['kustu] ['kuʃtu]	mineira, paulista, sulista etc. carioca, paraense, baiana etc.	antes de consoante desvozeada (surda)
	ASMA	['azma] ['aʒma]	mineira, paulista, sulista etc. carioca, paraense, baiana etc.	antes de consoante vozeada (sonora)
L	CALMA	['kauma] ['kauma] ['kaɬma] ['kaɹma]	carioca, mineiro, paulista etc. gaúcho de Porto Alegre gaúcho do interior "caipira"	qualquer contexto de travamento

Essa reduzida possibilidade de travamento silábico apresenta, além disso, uma característica muito interessante. Embora estejamos acostumados, na escola, a falar somente das vogais e das consoantes, existe uma terceira categoria de classificação dos segmentos sonoros, a das soantes. As vogais são produzidas sem nenhum tipo de obstrução, apenas com maior ou menor fechamento da boca. As consoantes, como nos informa a própria palavra, não podem soar sozinhas, é preciso que haja outro segmento para elas *soarem com* ele. As soantes, por seu turno, se situam entre as vogais e as consoantes, porque *soam* sozinhas, sem precisar do apoio de outros segmentos — são articulações nas quais ocorre somente um fechamento parcial ou então uma saída de ar desimpedida, oral ou nasal, o que as aproxima bastante das vogais. Ora, os segmentos chamados soantes englobam precisamente as líquidas ([r] e [l]), as nasais ([m] e [n]), as semiconsoantes ([ɪ] e [ʊ]), além de todas as vogais.

Essa característica das soantes permite que, em diversas línguas, elas ocupem o núcleo silábico, em lugar das vogais. As línguas eslavas (russo, polonês, tcheco, eslovaco, búlgaro, servo-croata etc.) se caracterizam por esse tipo de sílaba. À primeira vista, parecem palavras sem vogais, mas na verdade o núcleo da sílaba é ocupado por uma soante. Um exemplo muito famoso é a seguinte frase do tcheco:

Strč prst skrz krk
('Ponha o dedo na garganta' ou, literalmente, 'introduza o dedo através da garganta').

Observe que todas as palavras apresentam um [r] em seu centro.

Em inglês também é possível encontrar soantes como núcleo de sílaba. Observe as palavras abaixo e sua transcrição fonética:

- *battle* ['bætl] ('batalha')
- *castle* ['kæsl] ('castelo')
- *family* ['fæmli] ('família')
- *garden* ['gardn] ('jardim')
- *listen* ['lĭsn] ('ouvir')
- *little* ['lĭtl] ('pequeno')
- *world* ['wərld] ('mundo')

Em todas elas aparece uma soante [l], [r] ou [n] ocupando centro de sílaba. Essa qualidade das soantes também permite que elas se combinem em muitas palavras do inglês como *girl*, *gnarl*, *snarl*, *forlorn*, *worn*, *worm*, *squirm* etc.

Conhecendo as soantes e sua natureza quase vocálica, podemos dizer que no PB a maioria das sílabas ditas travadas têm, na verdade, como fator de travamento, não uma consoante propriamente dita, mas sim uma soante.

O [s], por seu turno, é uma consoante dita contínua, porque na sua produção a passagem de ar não se fecha completamente, de modo que sua emissão pode se prolongar por mais tempo, ao contrário das oclusivas, que fecham totalmente a passagem de ar. As outras contínuas são [f] e [v], [ʃ] e [ʒ], [z]. É isso que permite a [s], [ʒ] e [ʃ] ocupar a margem direita da sílaba, como vimos acima.

Convém ressaltar que em palavras como *pacto*, *rapto*, *istmo*, *ritmo*, *ficção*, *confecção*, *naftalina*, *pneu*, *advogado*, *submisso*, *magma*, *absurdo*, *atmosfera* etc., os falantes introduzem regularmente um [i] ou [e] epentético depois da consoante oclusiva, de modo a conduzir a sílaba ao padrão CV:

- *admitir* [adʒimi'tih]
- *advogado* [adʒivo'gadu] ~ [adevo'gadu]
- *afta* ['afɪta]
- *apneia* [api'nɛɪa]
- *atmosfera* [atʃimos'fɛra]
- *signo* ['siginu]
- *lactose* [laki'tɔzi]
- *pneu* [pin'eʊ] ~ [pe'neʊ]
- *réptil* ['hɛpitʃiɔ]
- *subsolo* [subi'sɔlu]

Observe que o [i] é tão nitidamente pronunciado que ele palataliza o [d] e o [t] que o precedem, exatamente como se dá com *dia* ['dʒia] e ['tʃia]. A epêntese do [e] caracteriza as variedades paulistana e sulistas de modo mais geral. Já durante o Romantismo, alguns poetas, tentando se valer da pronúncia brasileira, contavam seus versos considerando o [i] epentético:

"Ninguém mais *observa* o tratado" (9 sílabas métricas) - Gonçalves Dias
"Contudo os olhos d'*ignóbil* pranto" (10 sílabas métricas) - Gonçalves Dias

Esse recurso também está amplamente representado na música popular brasileira:

"Deus deu *resignação*
ao meu pobre coração" (7 sílabas métricas)
["Esta melodia", Bubu da Portela e Jamelão]

"Nem tampouco se *admite*
quem do nosso amor duvide" (7 sílabas métricas)
["Madalena", Ivan Lins e Ronaldo Monteiro de Sousa]

Conclusão: as chamadas consoantes "mudas" não existem no PB.

As mesmas soantes e contínuas que ocupam o final de sílaba também são as únicas permitidas em final de palavra no PB. Nessa posição, no entanto, mesmo sendo tão próximas das vogais, elas sofrem a pressão da "sílaba ideal", CV, de modo a permitir que toda palavra termine em vogal. Vamos ver como isso acontece.

O [r] final sofre apócope, isto é, desaparece. Essa supressão ocorre não só nas VLE como também nas VUP, em contextos de interação menos monitorada. No caso dos infinitivos verbais, a supressão do [r] já se tornou uma regra categórica — uma regra seguida por todos os falantes, independente de sua origem social ou grau de escolarização. Somente em situações de extremo monitoramento (como no caso dos locutores de rádio e televisão) o [r] é pronunciado, e mesmo assim de forma bastante branda:

- *cantar* [kɐ̃'ta] ~ [kɐ̃'taʰ]
- *beber* [be'be] ~ [be'beʰ]
- *dormir* [duɦi'mi] ~ [duɦimiʰ] etc.

Em outras palavras terminadas em [r], a supressão ocorre principalmente nas VLE ou nas interações menos monitoradas:

- *doutor* [do'to]
- *professor* [profe'so]
- *amor* [a'mo] etc.

No caso de [l], tanto em final de sílaba quanto em final de palavra, na ampla maioria das variedades linguísticas do PB, ocorreu a total vocalização dessa soante quando pós-vocálica. O que se escreve, portanto, vogal + l soa, de fato, como um ditongo, pronúncia considerada padrão para o rádio e a televisão:

- *anel* [a'nɛʊ]
- *Brasil* [bɾa'ziʊ]
- *farol* [fa'ɾɔʊ]
- *revolto* [he'voʊtu]
- *sinal* [si'naʊ]

Em algumas variedades rurais, ocorre uma paragoge, isto é, o acréscimo de uma vogal [i] após o [l]:

- *sol* ['sɔli]
- *papel* [pa'pɛli]
- *mal* ['mali]

No caso do [s] em final de palavra, a situação é bastante peculiar. Sendo, de fato, um morfema de plural, seria de se esperar que essa consoante não fosse eliminada. No entanto, por força da economia linguística [▶147] e da pressão do padrão silábico CV.CV, a morfologia de plural se preserva nos determinantes (artigo, possessivo, demonstrativo, quantificadores etc.), fazendo com que a pluralização se dê no plano do sintagma e não mais do paradigma [▶448]:

- os menino bonito
- esses carro bacana
- meus pacote
- cinco pedaço

O [s] também é frequentemente eliminado nas formas verbais da 1ª pessoa do plural, mesmo nas VUP, em situações distensas:

- *vamos* ['vãmu]
- *fizemos* [fi'zẽmu]
- *dissemos* [dʒi'sẽmu] etc.

O ditongo nasal *-ão*, se antecedido de *r-*, se conserva no final das palavras quando ocupa a sílaba tônica, ou seja, em palavras oxítonas. Quando, no entanto, ele se encontra em posição postônica, em palavra paroxítonas, portanto, a tendência dos falantes do PB é reduzi-lo à vogal [u], o que se verifica nas formas verbais do pretérito perfeito (não-pessoa do plural):

- *casarão* [kaza'ɾɐ̃ʊ̃] : : *casaram* [ka'zaru]
- *amarraram* [ama'haru]
- *comeram* [ku'meru]
- *fizeram* [fi'zɛru]
- *começaram* [kume'saru]

A respeito do apagamento das consoantes em final de palavra — que os puristas de língua inglesa classificam de *sloppiness* ("desleixo", "desmazelo", "falta de cuidado" etc.) —, Jean Aitchison, depois de passar em revista diversos exemplos, afirma:

> Esse desenvolvimento não é apenas "desmazelo", mas se deve à debilidade geral e inevitável da articulação dos sons no final das palavras [...].

> Tudo somado, portanto, é *normal* que as consoantes desapareçam nos finais das palavras no correr das eras. Já aconteceu em inúmeras línguas ao longo dos séculos e sem dúvida acontecerá em muitas outras. Que as palavras gradualmente percam suas terminações é um crime semelhante ao dos rios que erodem gradualmente seus leitos ou ao da chuva que desgasta o calcário (Aitchison, 2001: 155, 156).

Conforme já vimos, o tupi, o banto e o iorubá são línguas de padrão CVCV que desconhecem os encontros consonantais. Autores como E. Gärtner e Yeda Pessoa de Castro [▸239] atribuem ao contato dessas línguas com o português colonial a tendência geral do PB de buscar o padrão CVCV bem como, em muitas variedades rurais e/ou rurbanas, a eliminação dos encontros consonantais por meio de suarabácti:

- *caraca* (< craca)
- *convulussão* (< *convulsão*)
- *dificulidade* (< dificuldade)
- *fulô* (< flor)
- *Gulória* (< Glória)
- *parantá* (< plantar)
- *paricero* (< parceiro)
- *Quelemente* (< Clemente)
- *saravá* (< salvar)
- *terém* (< trem)

Tomando como base as variedades urbanas de prestígio, o inventário das sílabas possíveis no PB é o seguinte:

SÍLABA	POSIÇÃO NA PALAVRA		
	INICIAL	MEDIAL	FINAL
V	a.ba.ca.te	su.e.ca	ba.ú
VC	as.tro	a.or.ta	ca.ir
CV	ra.paz	co.bi.ça	la.men.to
CCV	pri.ma.ve.ra	a.cli.ma.ta.do	pe.dra
CVC	cer.to	ex.por.tar	va.mos
CVCC	pers.pec.ti.va	su.pers.ti.ção	
CCVC	pros.trar	sa.cris.tão	mi.la.gres
VV	eu.ro.peu	pi.au.i.en.se	cri.ei
VVC	eis	ma.es.tro [ma'ɛiʃtru]*	a.re.ais
CVV	pau.ta	re.jei.ção	de.co.rei
CCVV	trau.ma	in.troi.to	dis.trai
CVVC	cáus.ti.co	ho.lo.caus.to	de.pois
CCVVC	claus.tro	en.claus.trar	des.tróis

(*) Pronúncia carioca

7.11 O SOTAQUE E SEU PAPEL CULTURAL

O termo sotaque, que só existe em português e tem etimologia desconhecida, corresponde ao que em outras línguas se designa por *acento* (*accent* em inglês e francês; *acento* em espanhol; *accento* em italiano; *Akzent*, em alemão), termo técnico também empregado em português.

A palavra acento deriva do latim *accentu-* e esse, por sua vez, é resultante da junção da preposição *ad* e da palavra *cantu-* ('canto'). Por efeito da apofonia [▶296], o *a* tônico de *cantu-* se fechou em *e*. O *acento*, portanto, é a maneira como cada falante "canta" a sua língua, de acordo com a "melodia" própria de sua região, de sua classe social ou de sua etnia. É por isso que, ao falarmos do *sotaque* de alguém ou de algum grupo, é comum usarmos os adjetivos "cantado", "cantarolado" e semelhantes para designar o traço prosódico característico dessas pessoas.

Os sotaques se caracterizam por traços segmentais e por traços suprassegmentais. Os traços segmentais se referem a pronúncias características dos sons da língua: por exemplo, o "s chiado" dos cariocas, o "r caipira", as vogais pretônicas abertas dos falares nordestinos são traços segmentais. Os traços suprassegmentais têm a ver com a prosódia, isto é, com fatores como entoação, duração, linha melódica etc. É muito comum as pessoas usarem metáforas como "arrastado", "cantado", "ríspido", "suave", "bruto" etc. para se referir aos sotaques.

Os sotaques são as manifestações mais imediatas da identidade linguística dos falantes. Ao abrir a boca para falar, todo e qualquer falante, de toda e qualquer língua do mundo, exibe os traços prosódicos característicos de sua variedade linguística, de sua região, de sua classe social etc. São poucas as pessoas que conseguem, por uma aptidão natural ou por esforço consciente, "falar sem sotaque" — uma expressão que, embora muito usada, é uma contradição em termos, e vamos ver por quê.

Quando dizemos que alguém, ao falar sua língua (a questão das línguas estrangeiras é diferente), fala "sem sotaque", o que estamos realmente querendo dizer é que essa pessoa fala de um modo mais próximo do sotaque que, por razões exclusivamente históricas e socioculturais, se transformou numa espécie de "fala neutra" da língua nacional. O exemplo mais famoso é o do inglês britânico e de sua chamada *Received Pronunciation* (RP), que também recebe os pomposos nomes de *Queen's English* ("Inglês da Rainha"), *BBC English* ("inglês da rádio BBC"), *Oxford English* ("Inglês da Universidade de Oxford"). Essa pronúncia padrão se baseia nos modos de falar das classes sociais mais prestigiadas da região sul da Inglaterra, onde, não por coincidência, ficam a capital Londres e a cidade de Oxford, sede da famosíssima universidade, tão antiga que não se sabe exatamente quando foi fundada (os registros mais antigos remontam ao ano de 1096). Assim, na Inglaterra, "falar sem sotaque" significa, de fato, falar com o sotaque das classes altas do Sul da ilha.

No Brasil, como se sabe, a pronúncia padrão é aquela resultante de uma síntese das falas das camadas mais letradas das grandes cidades da região Sudeste, depois de eliminados os traços segmentais mais característicos. É a pronúncia que se emprega no *Jornal Nacional*, da Rede Globo, que é até hoje um dos programas de televisão mais assistidos do país. O apresentador William Bonner é o locutor-símbolo dessa pronúncia. Pesquisas realizadas por linguistas em todo o Brasil revelaram que uma grande maioria de brasileiros, ao serem indagados de como gostariam de falar, responderam: "Como os locutores do *Jornal Nacional*".

Essa pronúncia padrão assume um caráter "neutro" devido aos seguintes aspectos, entre outros:

- não apresenta o "s chiado" característico do Rio de Janeiro; em lugar dele, ocorre o [s] "sibilado";
- não apresenta o [ɾ] simples vibrado em sílaba travada, característico da fala paulistana, nem o [x] uvular de alguns falares cariocas; em lugar deles se usa uma aspiração leve, [h];
- apresenta as consoantes [ʤ] e [ʧ] diante de [i], características das variedades carioca, paulistana e belo-horizontina; isso porque a pronúncia não palatalizada dessas consoantes remete imediatamente a falares nordestinos;
- apresenta a semivogal [ʊ] na pronúncia do *-l* final de sílaba e de palavra, que também é comum às variedades carioca, paulistana e belo-horizontina;
- evita o fechamento das vogais pretônicas [e] e [o] em [i] e [u], como em *m̲enino*, *c̲omeço*, *p̲edir*, *m̲oeda* etc., pronunciando as vogais de um modo mais próximo do que está escrito.

Esse modelo de pronúncia é imitado, com maior ou menor sucesso, por locutores de rádio e televisão de outras regiões do Brasil. Mais uma vez, tal como no caso do inglês britânico, falar PB "sem sotaque" é falar como alguém das classes urbanas mais letradas da região Sudeste.

Na verdade, essa pronúncia padrão sintética se aproxima em grande medida, no que diz respeito a seus traços fonéticos (mas não prosódicos), da fala mineira mais característica da região de Belo Horizonte. Isso porque, historicamente, essa variedade já representa uma síntese de diversos falares brasileiros. Quando, no final do século XVII e início do XVIII, foram descobertas as ricas jazidas de ouro da região que, justamente por isso, passou a se chamar Minas Gerais (até então o território fazia parte da província de São Paulo, que se estendia até o Mato Grosso atual), ocorreu uma verdadeira "corrida do ouro" dirigida para lá. Vieram brasileiros de todas as demais regiões, na esperança de enriquecer com a exploração do minério precioso e, junto com eles, uma verdadeira multidão de escravos negros, trazidos principalmente das

lavouras açucareiras do Nordeste, então em declínio. Esses escravos é que seriam os escavadores das minas. Além disso, um grande contingente de portugueses, cerca de trezentos mil ao longo do século, também vieram da Europa em busca de riqueza. Nesse processo, o contato linguístico sem dúvida foi intenso entre (1) os brasileiros, que já falavam suas modalidades de PB e que decerto, sobretudo os (2) paulistas, ainda empregavam a língua-geral de base tupi, (3) os escravos, que falavam suas línguas ancestrais e também o PB resultante dos contatos entre essas línguas e o português colonial, e (4) os portugueses, que, no século XVIII, sem dúvida já falavam a modalidade de português europeu resultante das grandes transformações ocorridas na língua na virada do século XVII para o XVIII. A variedade mineira dessa região aurífera é, talvez, a que melhor representa uma síntese das muitas correntes linguísticas que confluíram para a formação do PB mais característico: a indígena, a africana, a brasileira e a portuguesa.

Os sotaques são resultantes de diversos processos. No PB, por exemplo, as variedades nordestinas e, mais ainda, as das regiões mais interioranas (sertão), apresentam traços mais conservadores, com diversos vestígios do que deve ter sido a língua trazida pelos portugueses nas primeiras fases do período colonial. Isso se verifica não somente na fonética e no sotaque, mas também no léxico e na sintaxe. Os nordestinos fazem, por exemplo, em seu falar corriqueiro, usos do modo subjuntivo que para os habitantes do Sudeste soam estranhos, mas que aparecem muito bem registrados na literatura medieval e clássica: "Não sei o que *faça* para esse menino comer direito".

O sotaque carioca, sem dúvida, deve muitos de seus aspectos à transferência da Corte portuguesa para o Rio de Janeiro em 1808, um aporte populacional de cerca de 15.000 pessoas, cujo modo de falar, prestigiado, decerto passou a ser imitado pelos demais habitantes da cidade. Em todas as demais zonas costeiras do Brasil onde impera o "s chiado", é possível retraçar uma presença significativa de portugueses vindos de Portugal ou das ilhas dos Açores.

O sotaque paulistano tradicional decorre do contato de falantes de diversos dialetos italianos com o PB. Isso explica alguns traços característicos da variedade paulistana, como a desnasalização das vogais tônicas de *homem, come, Antônio, viemos*, que ali soam abertas ("hòmem", "còme", "Antònio", "vièmos"). Esse sotaque, no entanto, está se tornando minoritário, à medida que a população de São Paulo se diversifica, com uma forte imigração nordestina e de paulistas do interior, além de muitos estrangeiros (coreanos, chineses, bolivianos, ao lado dos japoneses, estabelecidos há mais tempo).

No interior dos estados da região Sul, existem diversas comunidades em que o PB não é a primeira língua aprendida pelos habitantes e sim algum dialeto alemão, com predomínio do *hunsrückisch*, empregado no ambiente familiar de um número considerável de brasileiros. Daí decorre, por exemplo, a pronúncia [ɾ] tanto para *caro* quanto para *carro*, uma vez que não existe distinção entre duas consoantes vibrantes

em alemão. Também é comum nessas variedades a pronúncia surda [ʃ] para o que se escreve com *ge*, *gi* ou *j*, pois o dialeto também não conhece a consoante sonora palatal [ʒ]. Esses falantes sofrem, por isso, muita discriminação quando confrontados aos falares das grandes cidades da região.

Na região de Cuiabá se desenvolveu uma variedade com traços muito peculiares devidos, ao que parece, a um substrato indígena bororo [▶326].

Em Brasília, após os primeiros cinquenta anos de fundação da capital, os estudiosos começam a detectar alguns traços capazes de configurar um futuro sotaque brasiliense. A influência nordestina sem dúvida é a mais marcante, ao lado da mineira e da carioca, bem como da goiana, uma vez que o Distrito Federal se localiza dentro do estado de Goiás.

O sotaque exerce um forte papel nas avaliações socioculturais que os falantes da língua recebem em suas interações sociocomunicativas. Devido às condições socioeconômicas de determinadas regiões, as falas características desses lugares receberão avaliação positiva ou negativa de acordo com o prestígio político e/ou a importância econômica da região. Isso explica a forte carga de desprestígio que pesa sobre as variedades nordestinas, identificadas com uma região tida como "atrasada" politicamente e "subdesenvolvida" economicamente. Um falante nordestino no Sudeste é facilmente identificado por seu sotaque e, em decorrência disso, pode vir a sofrer muita discriminação. Prova disso é o emprego recorrente de personagens nordestinos em novelas de televisão, sempre interpretados por atores e atrizes não nordestinos que falam com um sotaque estereotipado e exagerado, com o fim de obter efeitos de humor e depreciação.

A autodepreciação também é um fenômeno corrente, provocado pela depreciação que determinados sotaques sofrem quando confrontados com outros, mais prestigiados. Assim, muitos nordestinos confessam não gostar de seu modo típico de falar a língua e procuram adquirir os traços segmentais e suprassegmentais das variedades de maior prestígio. Falantes do "dialeto caipira", quando vão para as grandes cidades, principalmente São Paulo, tentam apagar os traços mais característicos de suas variedades linguísticas.

Na prática pedagógica, é muito importante que os docentes estejam atentos aos mecanismos de discriminação que podem ser ativados com base nos sotaques de seus alunos. Principalmente quando a sala de aula é mais homogênea, com alunos nascidos num mesmo lugar, é muito comum ocorrer atitudes de zombaria diante de alunos provenientes de outras regiões. Uma escola democrática e democratizadora tem de respeitar a diversidade linguística e impor esse respeito na formação de seus alunos.

a Manoel Luiz Gonçalves Corrêa

8

ruídos & rabiscos

língua falada e língua escrita

8.1 A FALÁCIA CLÁSSICA

Para entendermos bem o que está implicado nas relações entre língua falada e língua escrita, temos de convocar os fundadores da gramática tradicional, os filólogos alexandrinos [▶420]. O principal interesse desses funcionários da Biblioteca de Alexandria era chegar a uma versão o mais fiel possível dos grandes poemas homéricos,

a *Ilíada* e a *Odisseia*, símbolos maiores da cultura e da civilização gregas. Para empreender esse trabalho, eles criaram um aparato terminológico e critérios de classificação, exatamente como faz qualquer pesquisador, cientista ou teórico dedicado ao estudo de um objeto específico. Por isso é que cada área de conhecimento — física, química, botânica, zoologia, astronomia, geografia, história, filosofia etc. — tem sua terminologia própria e suas próprias classes, categorias, tipos, gêneros, espécies e por aí vai. A taxonomia — a teoria e a prática da classificação — deve ser sempre vista, portanto, como um *meio*, um *método*, um *instrumento* para a melhor compreensão das coisas e dos eventos, e não como um *fim* em si mesma. E precisamente por isso é que as taxonomias (palavra grega que significa "regras de organização") são regularmente revistas, refeitas, ampliadas, reduzidas, de acordo com o ritmo do progresso de cada área de conhecimento.

Um exemplo recente que chamou muito a atenção foi a decisão da União Astronômica Internacional, tomada numa assembleia geral da entidade em 2006, de retirar Plutão da lista dos planetas do nosso sistema solar e incluí-lo num grupo de objetos celestiais denominados, em inglês, KBO (Objetos do Cinturão de Kuiper — uma área do sistema solar que fica para além dos planetas). Isso se deveu, entre outras coisas, à descoberta, em 2005, de um corpo celeste mais volumoso que Plutão, que foi batizado de Éris, nome da deusa grega da discórdia. Se Plutão tivesse de continuar sendo considerado como um planeta, seria preciso incluir Éris também nessa classificação. Em vez disso, os astrônomos preferiram reclassificar Plutão, e sua decisão foi acatada por ter sido tomada pelo órgão máximo da astronomia mundial.

Essa é uma das razões por que a maioria dos linguistas se opõem à existência de uma nomenclatura oficial para os estudos gramaticais (a NGB, em vigor no Brasil há mais de cinquenta anos) e, principalmente, para o ensino de língua materna. Vamos tratar disso mais explicitamente em [▸502].

Também é por isso que, nessa gramática, me concedo ampla liberdade para classificar os fenômenos linguísticos de maneira às vezes radicalmente diferente das classificações tradicionais. O importante, como tudo o que diz respeito à construção do conhecimento, não é encontrar a classificação *mais certa* ou mais próxima da "verdade" — como em toda investigação científica, técnica, filosófica etc. O importante é oferecer argumentos coerentes que justifiquem a adoção dessa ou daquela classificação em lugar de outras.

A doutrina gramatical tradicional está marcada, portanto, desde suas origens, por um apego exclusivo à língua escrita e, mais restritamente ainda, à língua escrita literária. Isso terá profundas consequências no futuro dos estudos gramaticais pelos próximos dois mil e trezentos anos.

8.1.1 Dois pressupostos falsos

O edifício da doutrina gramatical tradicional foi erguido sobre dois alicerces frouxos e cambaleantes que o linguista John Lyons (1968: 9) rotulou de "a falácia clássica". São dois pressupostos falsos. Um deles diz respeito à mudança linguística: tomando consciência de que a língua falada em seu tempo era diferente da língua escrita registrada nos textos da grande literatura grega do passado, os primeiros gramáticos não só deduziram o óbvio, que a língua tinha mudado, como também lançaram um juízo de valor negativo sobre essa mudança — a língua tinha mudado *para pior*, o seu estado atual representava uma *ruína*, uma *decadência*, uma *corrupção* da língua "pura" e "perfeita" usada pelos grandes autores muitos séculos antes.

O segundo pressuposto errôneo, na verdade o outro lado dessa moeda falsa, é o de que a língua falada era caótica e desregrada, não tinha organização gramatical e, por isso, para que alguém falasse de um modo minimamente "elegante" e "correto", era preciso se basear nas regras preconizadas para a escrita literária.

Os gramáticos alexandrinos, ao incorrer nessa "falácia clássica", também lançaram as sementes do preconceito linguístico, conjunto de mitos e superstições sobre a(s) língua(s) e seus falantes, que frutificou com fartura por toda a história do Ocidente, chegando praticamente intacto aos dias de hoje, quando a retumbante maioria das pessoas acredita piamente que existe um modo "certo" e um modo "errado" de falar a língua, que as pessoas sem instrução formal "falam tudo errado", que existe "algum lugar no passado" onde a língua foi falada e escrita "corretamente", que a língua falada não tem regras e que "só a escrita tem gramática" e por aí vai...

Outra consequência perniciosa da falácia clássica é a polarização maniqueísta, como classificou o filólogo Antônio Houaiss (1985: 133). Separadas como entidades monolíticas e estanques, como dois polos muito distantes entre si, a fala e a escrita também são analisadas no senso comum com base na dicotomia positivo/negativo, razão do adjetivo *maniqueísta* usado por Houaiss. Evidentemente, o polo positivo é reservado à escrita e o negativo, à fala. Daí a frequente alegação de que é preciso eliminar "as interferências da fala na escrita", como se tal quimera fosse possível. Afinal, se a escrita foi inventada como um meio de registrar a fala, eliminar as interferências da fala na escrita seria, pura e simplesmente, eliminar a escrita.

Essa "língua purificada", ou seja, uma "escrita purificada", não existe nem poderia existir porque a heterogeneidade é constitutiva das línguas humanas. Não existe um padrão predeterminado, compacto, homogêneo e perfeito de "escrita" porque seus contornos variam de falante para falante (e até mesmo de gramático para gramático). A concepção de "correção" de cada falante varia muito: tem aqueles que acham que é preciso "modernizar" a língua e admitir o que "todo mundo fala"; outros conside-

ram que esse "liberalismo" tem de ser limitado; outros, por fim, condenam qualquer coisa que não esteja prevista na norma-padrão canonizada.

Os gramáticos da Antiguidade cometeram, portanto, dois erros fundamentais, ambos derivados da constatação de fatos óbvios aos quais, no entanto, eles atribuíram valores morais negativos, mas coerentes com sua visão de mundo e seus próprios preconceitos sociais e culturais. Inadmissível é querer, em pleno século XXI, continuar propagando essa ideologia linguística preconceituosa, depois de tantos avanços científicos e de tantas conquistas positivas no campo dos direitos humanos.

A falácia clássica reside na comparação entre duas entidades sociolinguísticas completamente distintas:

(i) a língua espontânea falada na Alexandria do século III a.C.; e
(ii) a língua escrita literária da Atenas do século V a.C.

Temos aqui uma análise distorcida, resultante de três confusões:

(1ª) a confusão de registro de modalidade (fala espontânea/escrita monitorada);
(2ª) a confusão temporal (século VIII-V a.C./século III a.C.);
(3ª) a confusão espacial (Atenas, Grécia, Europa/Alexandria, Egito, África).

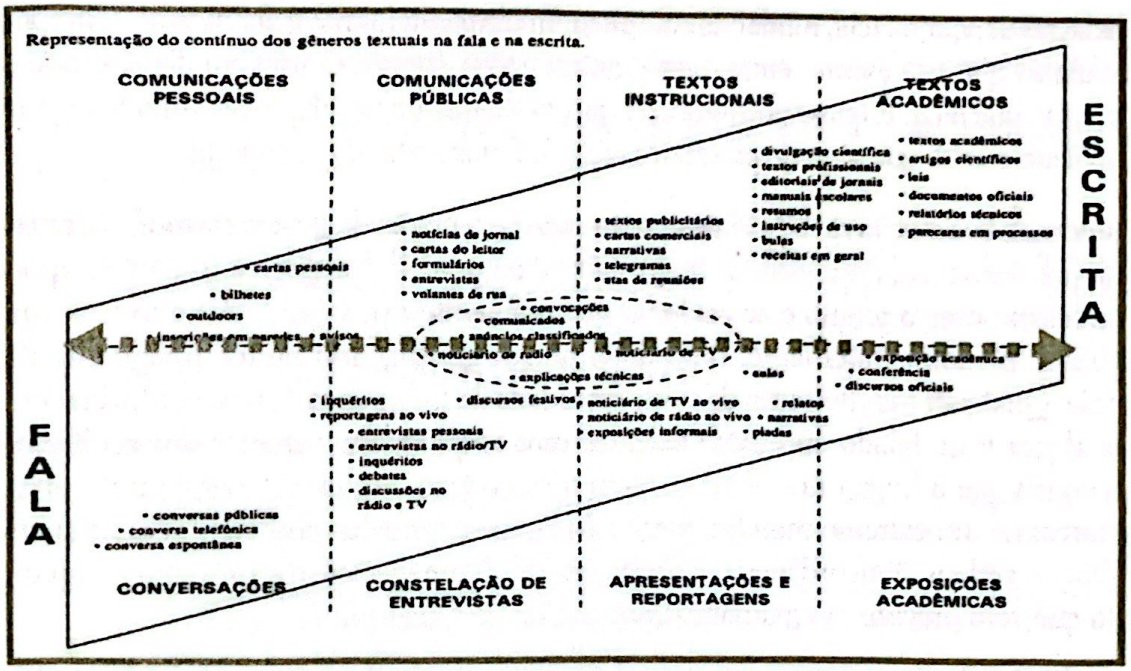
Ora, não se pode confundir a fala espontânea com a escrita literária monitorada. Não estamos aqui defendendo a rígida separação entre fala e escrita (como fizeram, equivocadamente, os filólogos alexandrinos e que a maioria das pessoas ainda hoje acredita existir), mas sim a impossibilidade de comparar *registros, estilos de uso da língua* completamente diferentes, bem como diferentes *gêneros discursivos*.

Entre língua falada e língua escrita existem muito mais semelhanças do que diferenças. De fato, essas duas modalidades de uso da língua só diferem quanto às condições de produção: a fala é produzida e revisada simultaneamente, *on-line*, enquanto a escrita permite que produção e revisão ocorram em momentos distintos.

8.1.2 O hibridismo inevitável

O linguista brasileiro Luiz Antônio Marcuschi, pioneiro dos estudos sobre língua falada entre nós, postula a existência de um espectro contínuo que vai do mais falado para o mais escrito, espectro atravessado pela variação estilística, que vai do gênero textual menos monitorado para o gênero textual mais monitorado (lembrando que a palavra *texto* se aplica tanto ao que se fala quanto ao que se escreve), espectro que ele apresenta sob a forma do diagrama a seguir (2001: 41):

Como se vê, existe fala espontânea e escrita espontânea, como também existe fala formal e escrita formal. A seta tracejada (que acrescentei ao diagrama original de Marcuschi) representa a *variação estilística*, do menos monitorado ao mais monitorado. Se for para empreender algum tipo de comparação, ela tem que ser feita entre



registros de uso semelhantes (gênero textual falado espontâneo e gênero textual escrito espontâneo, por exemplo) e nunca entre os dois extremos do espectro fala-escrita e do *continuum* de monitoramento estilístico. Entre textos *falados* formais e textos *escritos* formais, por exemplo, as semelhanças são muito mais consistentes e evidentes do que as diferenças, o que impossibilita a manutenção da separação rígida entre fala e escrita que desconsidera a *variação estilística* e a variação de *gênero textual*.

A contribuição de Marcuschi sem dúvida tem sido fundamental para aprimorar nosso entendimento das relações entre língua falada e língua escrita. No entanto, desde que sua proposta de análise foi lançada, ocorreram transformações aceleradas nos modos e nas condições de produção da fala e da escrita, transformações ocasionadas pelo desenvolvimento vertiginoso dos novos meios de comunicação e das novas tecnologias de informação.

Nem mesmo as diferenças devidas ao gênero textual e à variação estilística podem servir hoje de argumento para qualquer tipo de distinção nítida entre fala e escrita. Toda produção textual na atualidade, falada e/ou escrita, se configura inexoravelmente como uma manifestação semioticamente híbrida que mobiliza os multimeios sonoros, visuais, gráficos, tridimensionais etc. que as novas tecnologias de comunicação e informação têm colocado ao nosso dispor.

Outro aspecto que cabe ressaltar é que essas mesmas novas tecnologias possibilitaram que a escrita dispusesse de recursos semióticos amplamente capazes de concorrer com as manifestações orais. Na fala, podemos nos valer do tom de voz, da altura, das infindáveis modulações de entoação, do falsete, da imitação de outras vozes ou de vozes de animais, da gagueira proposital, do arremedo de sotaques, da ênfase por meio da separação das sílabas, entre tantos outros recursos. Até pouco tempo, quando comparada a essa riqueza de possibilidades, a escrita parecia muito mais pobre. Hoje, no entanto, temos à nossa disposição sofisticados recursos de multimídia que nos permitem aumentar o tamanho das letras para enfatizar a mensagem, usar cores variadas, fazer as

palavras dançar na tela, mudar de cor, piscar intermitentemente e até mesmo sobrepor a fala ao que está escrito, entre muitas outras coisas. Escrever, hoje em dia, é quase o mesmo que falar. É quase possível dizer que, tanto quanto na fala, existe uma prosódia no texto escrito que se pode produzir graças às ferramentas da era digital.

Retomando a noção de hibridismo, nem mesmo a noção de *gênero textual*, recentemente introduzida nos estudos linguísticos e na prática pedagógica, pretende apreender um objeto pronto e acabado. O que realmente existe são *textos* que se configuram, predominantemente, num determinado *gênero*, mas nunca integralmente nele. Qualquer manifestação da nossa faculdade de linguagem é, repito, híbrida: em qualquer texto falado ou escrito fazemos usos amplamente variados dos múltiplos recursos que a língua nos oferece. Num mesmo texto em que encontramos certas marcas de um extremo monitoramento do discurso também podemos encontrar regências verbais, concordâncias, colocações pronominais e outros usos que escapam do que vem previsto nas gramáticas normativas. Por exemplo:

(1) As falhas operacionais na ocupação do Complexo do Alemão derivaram da falta de conhecimento do fenômeno e isso não significa acordo com o crime organizado. O diversionismo em curso só aproveita ao crime organizado. Ele quebra a confiança dos cidadãos nas forças do Estado. Durante anos assistiu-se, nas correlações entre associações criminosas e membros escravizados da comunidade, um vínculo de solidariedade constituído pelo medo. E tal vínculo acabou de ser desfeito com as retomadas. (Wálter Fanganiello Maierovitch, *Carta Capital*, nº 625, 8/12/2010, p. 31)

Observe que o autor escreve com claro domínio das convenções dos gêneros textuais mais monitorados, e a última coisa que se pode dizer desse trecho é que ele é "informal". No entanto, ao lado de uma regência clássica, de emprego bastante raro — "o diversionismo em curso só aproveita ao crime organizado" —, encontramos outra — "assistiu-se [...] um vínculo de solidariedade" —, em que ocorre ao mesmo tempo a *ênclise*, colocação pronominal contrária às preferências sintáticas do PB, e o uso do verbo *assistir* como transitivo direto, regência censurada indefectivelmente em todas as manifestações puristas e em todos os materiais didáticos filiados à TGP.

Distração? Lapso? Não, apenas o inevitável caráter *híbrido, heterogêneo* de qualquer produção textual, em qualquer gênero discursivo. E às vezes num mesmo trecho encontramos flutuações no uso, como no seguinte caso, com a regência do verbo *implicar*:

(2) Uma mudança no formato dos jornais implica obrigatoriamente em uma mudança de concepção. E dada a preferência que a nossa imprensa mimetizou da TV, implicará uma alteração de teor. (Alberto Dines, *Jornal do Brasil*, 10/7/1999)

Da mesma forma, um falante altamente letrado, em situação de grande monitoramento de sua fala e/ou escrita, pode ser levado, pelo desejo de "acertar" sempre, a produzir hipercorreções, isto é, a aplicar as regras normatizadas mesmo em situações onde elas não são previstas (ou são mesmo condenadas) pela TGP:

(3) É a economia dos novos tempos, que, apontam diversos especialistas, não há motivo para não dar certo. (Dal Marcondes, *Carta Capital*, nº 625, 8/12/2010, p. 55)

Provavelmente, devido à pressão prescritivista, que proíbe o uso do verbo *ter* com sentido existencial (*Tem muita gente aqui hoje*), o autor usou o verbo *haver* — no entanto,

aqui o verbo *ter* é o único adequado, já que admite sujeito ("a economia dos novos tempos que" ... "não *tem* motivo para não dar certo"), ao contrário do *haver* existencial.

Uma combinação de suposta "interferência da língua falada" e de hipercorreção ocorre no texto abaixo:

(4) Provavelmente receberá como resposta uma listagem de quase toda a informação disponível e, se não dispor de tempo para seleção e filtragem, provavelmente será esmagado pelo excesso de informação [...] Afinadíssimo com as mais recentes pesquisas arqueológicas e antropológicas, passa longe dos preconceitos de que não podem haver ideias dignas desse nome na mente dos "primitivos" [...] (Elias Thomé Saliba [historiador, professor da USP], *O Estado de S.Paulo*, 19/12/2004, caderno 2, p. D4).

Outro hibridismo muito comum ocorre no seguinte trecho:

(5) O novo governo basco promete coordenar esforços com a Catalunha para negociar a independência. Nesta região, esperam-se resultados parecidos na eleição de 25 de novembro: socialistas devem cair de 28 para 21 ou 22 cadeiras, o PP manter suas 18 ou perder uma e os separatistas e federalistas tomarem o restante das 135 cadeiras, com aumento de 62 para 68 a 69 para o partido CiU do atual governo, que promete um plebiscito pela independência. (CartaCapital, 31/10/2012, n. 721, p. 20)

Temos aqui a obediência à regra tradicional da chamada "voz passiva sintética" ou "passiva pronominal", que bate de frente com a intuição linguística dos falantes de PB: "esperam-se resultados", com o verbo no plural em concordância com seu objeto direto. Afinal, como se recitava no ensino tradicional, "quem espera espera alguma coisa", e o que se espera no caso são "resultados", ou seja, "resultados" é o óbvio objeto direto de esperar. O verbo no plural indica consciência de uma regra que, apesar de ilógica, vem preconizada em todos os manuais de redação. No entanto, o autor faz uma concordância que qualquer desses manuais condenaria ao encadear a sequência "devem cair... [devem] manter... [devem] *tomarem*". Essa flexão do infinitivo vem se tornando cada vez mais comum quando o verbo auxiliar da locução fica muito distante no encadeamento do texto e o autor sente necessidade de explicitar o plural do sujeito. É o mesmo que se encontra em:

(6) Poderíamos fazer como a crítica no passado, e falarmos da classicização da literatura contemporânea, por um lado, em um movimento que tende ao estéril, e da perda do mínimo fôlego de um estranho e mais do que retardatário beatniquismo [...] (Tales Ab'Sáber, *Lulismo: carisma pop e cultura anticrítica*, São Paulo, Hedra, 2011, p. 73).

Outro exemplo interessante de hibridismo é o do trecho abaixo:

(7) [...] libertar os judeus do confinamento daquela atividade comercial (que havia, segundo ele, se tornado a essência do judaísmo) [...] (Mary Gabriel, *Amor & Capital*, tradução de Alexandre Barbosa de Souza, Rio, Zahar, 2013, cap. 5, "Paris, 1843", e-book).

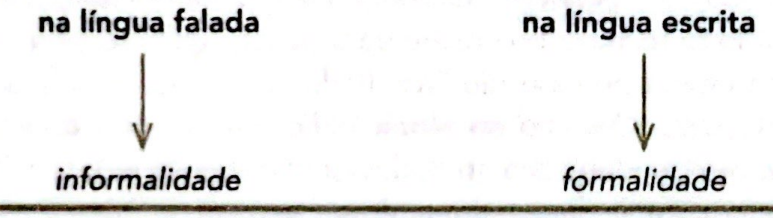
Mais de 90% dos pretéritos mais-que-perfeitos empregados nessa tradução são compostos com o auxiliar *haver*, como no exemplo acima, o que é uma claro indício de hipercorreção, de tentativa de "fugir da oralidade", já que a forma mais usual no PB é o mais-que-perfeito composto com o auxiliar *ter* ("tinha se tornado"). No entanto, ao lado dessa hipercorreção aparece um uso que, igualmente brasileiríssimo, é veementemente condenado pela tradição gramatical: a separação do pronome oblíquo do verbo auxiliar (*havia, segundo ele, se tornado*), quando o "certo", pela mesma tradição, seria "que se havia tornado, segundo ele, a essência do judaísmo").

Esse caráter aparentemente movediço e escorregadio da linguagem não deve nos assustar. Pelo contrário, ele prova o quanto a linguagem tem de *humano* e, portanto, de criativo, de inventivo, de indeterminado, de heterogêneo, de variável, de mutante, de opaco, de inconsciente, de falível, de inovador, de repetitivo, de compartilhável...

8.1.3 Outra polarização falaciosa

Ecoando a polarização maniqueísta, se cristalizou na cultura ocidental uma dupla de sinônimos, completamente falsa e equivocada: "fala = informalidade", "escrita = formalidade", equívoco que vem estampado, de maneira mais ou menos explícita, em vários dos livros didáticos disponíveis no mercado e adotados como material de ensino. Um exemplo é o seguinte:

Na língua falada, o emissor não tem a preocupação de usar as palavras com sentido preciso. Ele se comunica de uma maneira mais espontânea, mais familiar. Na língua escrita, ao contrário, há uma preocupação maior com as normas da língua. Ela deixa de ser informal, para se tornar mais formal.



[Sargentim, 2008, vol. 8: 15]

A simples oposição, tão frequente nesses materiais didáticos, entre "culto" e "coloquial" é um sintoma da manutenção dessa falácia: o termo *colóquio*, em todas as definições que os dicionários dão a ele, inclusive na esfera acadêmica, se refere, como aparece no *Houaiss*, a "conversa", "conversação", "debate". A oposição *culto/padrão* vs. *coloquial/oral* leva a crer, enganosamente, que não existe uma variedade culta informal escrita nem uma variedade culta formal falada, o que é uma rematada inverdade. Um exemplo dessa atitude (que também revela um pesado preconceito social):

O uso popular é sempre informal e quase sempre oral.
O uso culto pode ser formal ou informal, oral ou escrito.
Veja o esquema:

USOS DA LÍNGUA	
CULTO	FORMAL
	INFORMAL
POPULAR	

[Teixeira e Disciní, 2008: v. 7, p. 128; v. 8, p. 114.]

A escrita tomada em bloco, como se fosse sempre "formal", leva à equiparação equivocada de "linguagem popular" com "oralidade", como se não existisse língua "culta" oral:

> Na linguagem popular, oral, são comuns as ocorrências em que o verbo não concorda com o sujeito: *Nós já* ***fez*** *tudo o que* ***podia***. (Faraco e Moura, 2008, vol. 6: 84)

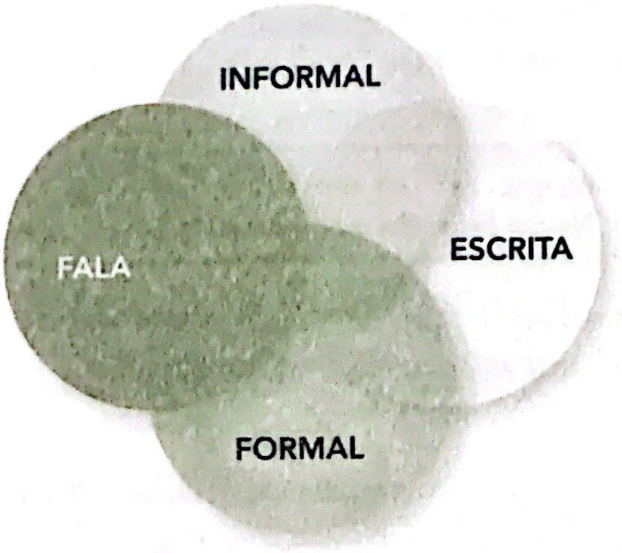
Ora, a concordância verbal (e nominal) variável também se verifica intensamente nas *normas urbanas de prestígio faladas menos monitoradas*. O exemplo é profundamente infeliz.

Outra consequência dessa visão deturpada da relação entre fala e escrita é a falsa condescendência dos propagadores da TGP, que alegam que "falar assim, tudo bem, mas escrever, nunca", numa suposta defesa da língua falada, sempre identificada com a "informalidade". No entanto, como vamos demonstrar fartamente nos próximos capítulos, muitas das formas que, nessa perspectiva, só deveriam aparecer na "fala" já comparecem nos textos escritos em gêneros mais monitorados, desmontando completamente a equiparação "fala = informalidade".

A velha e falaciosa equiparação — *escrita = formalidade / fala = informalidade* — se originou, como dissemos, do preconceito dos primeiros gramáticos contra a língua falada e de sua atitude de hipervalorização da escrita literária antiga, tomada como única forma "correta" de uso da língua. No entanto, essa associação, que já nasceu equivocada, se torna ainda mais descabida nos dias de hoje em que, graças aos avanços tecnológicos, é possível usar a escrita para a produção de textos efêmeros e altamente informais (como as mensagens enviadas por telefone celular), ao mesmo tempo em que é possível registrar de modo mais duradouro a produção oral mais monitorada (gravações em mídias de diversos tipos). Já existe mesmo toda uma produção literária de narrativas breves destinadas a serem lidas na tela do telefone celular ou dos chamados *tablets* (suportes para a leitura de livros, jornais, revistas etc. digitalizados).

Dessa forma, nossa época conhece a existência de gêneros híbridos, em que a separação entre *fala e escrita* se torna virtualmente impossível: os bate-papos via internet, por exemplo, apesar de se valerem da escrita, apresentam estrutura muito semelhante à de uma conversação oral (e o próprio nome *bate-papo* demonstra isso). A possibilidade, em fase de experimentação avançada, de ditar textos para serem escritos por um computador também aponta para essa dissolução das fronteiras entre oral e escrito.

Esses quatro termos, portanto, compõem uma rede complexa de possibilidades de uso, com múltiplas interpenetrações e interinfluências, e nunca dois universos estanques.



8.1.4 A *confusão de escrita com ortografia*

Além da redução do conceito de escrita somente à escrita literária (e, mesmo assim, dentro de uma concepção também restrita de literatura, que só contempla os autores devidamente canonizados sob o rótulo de "clássicos"), outro grave problema muito presente no ensino é a sinonímia equivocada *escrita = ortografia*. A tal ponto que uma elevada porcentagem do que se rotula, na tradição escolar, de "erro de português" é, na verdade, mero desvio com relação à ortografia oficial. O vigor desse equívoco pode ser depreendido, por exemplo, de um exercício de pesquisa sugerido por um livro didático. Depois de apresentar o já surrado exemplo do poema "Erro de português", de Oswald de Andrade (onipresente nos materiais didáticos), os autores pedem ao aluno:

> Procure localizar erros de português em cartazes, placas, ou até mesmo na fala das pessoas que você conhece. Transcreva-os em seu caderno. (Carvalho e Ribeiro, 1998: 125)

Ora, em cartazes e placas não aparecem "erros de português" e, sim, desvios com relação à ortografia oficial. Escrever, por exemplo, LOGINHA DE ARTEZANATO, onde a lei obriga a escrever LOJINHA DE ARTESANATO em nada vai prejudicar a intenção do autor da placa: informar que ali se vende objetos de artesanato. Nem sequer a realização fonética da placa "certa" e da placa "errada" vai apresentar diferença. O fato também de haver "erro" na placa não dá certeza nenhuma de que os objetos ali vendidos sejam de qualidade inferior, "errados" ou "feios".

Evidentemente, não se pode desprezar o valor simbólico que o erro ortográfico assume nas relações sociais, servindo muitas vezes como critério para discriminação e exclusão de uns e o favorecimento de outros. Quase sempre, um único desvio ortográfico basta para invalidar um texto completo e lançar sobre o autor a pecha de que "não sabe português".

Mais acima escrevi *lei* e não foi por acaso. A ortografia oficial é fruto de um gesto político, é determinada por decreto, é resultado de negociações e pressões de toda ordem (geopolíticas, econômicas, ideológicas). No início do século XX, o certo era escrever: NA PARAHYBA ELLE POUDE ESTUDAR AS SCIENCIAS NATURAES COMO A PHYSICA E A CHIMICA. Se hoje o certo é escrever NA PARAÍBA ELE PÔDE ESTUDAR AS CIÊNCIAS NATURAIS COMO A FÍSICA E A QUÍMICA, isso não altera a sintaxe nem a semântica do enunciado: o que mudou foi somente a ortografia.

Por isso, seria pedagogicamente mais proveitoso substituir a noção de "erro" pela de tentativa de acerto. Embora possa soar como simples eufemismo do tipo "politicamente correto", na verdade é possível obter um efeito significativo ao trocar um termo de conteúdo negativo (erro) por outro de conteúdo positivo (acerto). Afinal, a língua escrita é uma análise da língua falada, e essa análise será feita, pelo usuário da escrita no momento de grafar seu texto, em sintonia com seu perfil sociolinguístico.

Uma pessoa com poucos anos de escolarização, pouco habituada à prática da leitura e da escrita, isto é, com nível rudimentar de letramento e tendo como referência apenas uma suposta equivalência unívoca entre som e letra, fará uma análise dotada de reduzido instrumental teórico, empregando como ferramenta básica a *analogia*. Assim, quem escreveu CHINGAR em vez de XINGAR não fez isso porque quis errar, mas sim porque quis acertar. Se existe CHINÊS, CHIMARRÃO, CHICÓRIA, CHICOTE etc., por analogia se chega à possibilidade de também existir CHINGAR.

É fundamental ter em mente que os erros de ortografia são constantes, seguem uma lógica coerente e que, portanto, são facilmente previsíveis: são as trocas de J por G, de S intervocálico por Z, de CH por X, de Ç por SS, de L por U etc. e vice-versa, justamente por serem casos em que é necessário fazer uma análise que ultrapassa os limites teóricos da suposta equivalência som-letra. Dificilmente alguém vai tentar escrever XINGAR com J, G, S ou Ç no lugar do X oficial, porque faltam dados de experiência para uma analogia razoável.

Por outro lado, uma pessoa que tenha frequentado a escola por muitos anos, que leia e escreva assiduamente, que tenha se familiarizado com o uso dos dicionários, que tenha sido instruída sobre as regularidades e irregularidades da ortografia — ou seja, alguém com bom nível de letramento — vai saber que a simples analogia não é suficiente como guia no momento de escrever. Será preciso recorrer a outros quadros de referência: os gêneros textuais mais monitorados, a etimologia das palavras, os critérios de normativização da ortografia, tradições culturais etc.

No que diz respeito à língua falada, já se sabe que o rótulo de "erro" é aplicado a toda e qualquer manifestação linguística (fonética, morfológica e sintática, princi-

palmente) que se diferencie das regras prescritas pela tradição gramatical prescritiva, que se apresenta como codificação da "língua culta", embora na verdade seja a codificação de um padrão idealizado, que não coincide com nenhuma variedade urbana de prestígio real.

Segundo essa conceituação, são igualmente "errados" os dois enunciados abaixo:

(8) A Mercedes, ela é uma mulher que eu tenho muita admiração.
(9) *A Mercedes que eu tenho ela admiração é muita uma mulher.

Ora, o enunciado (8) é perfeitamente inteligível, decodificável, interpretável e, portanto, *gramatical*, aceitável por parte de qualquer falante nativo do PB. Por seu turno, (9) é nitidamente *agramatical* e, por isso mesmo, não ocorre na fala normal de nenhum brasileiro. No entanto, (8) é considerado tão "errado" quanto (9) pela TGP porque nenhum dos dois enunciados se enquadra nas prescrições tradicionais. O enunciado (8), contudo, tem propriedades sintáticas, semânticas e discursivas que qualquer falante de PB reconhece com tranquilidade — e a prova disso é que enunciados desse modelo são proferidos aos milhões diariamente em todos os cantos do país, por pessoas de todas as classes sociais e graus de letramento. O que encontramos nele é, de um lado, a construção *tópico-comentário* [▸472] e, do outro, a reanálise gramatical do pronome relativo *que* [▸919].

Diante desses fatos é que o antropólogo canadense Stephen Aléong propõe que a noção de erro linguístico fique reservada

> a fenômenos marginais ou prejudiciais à inteligibilidade da comunicação. A partir do momento em que um uso ou, melhor, uma regra ultrapassa o plano individual e conhece uma frequência de emprego para além de um nível determinável, trata-se já de uma regra a ser inscrita no repertório linguístico da comunidade. O erro, portanto, não é absoluto mas, sim, relativo ao meio ou ao grupo social de referência. Segundo nossa definição, o erro é essencialmente um uso que, num dado momento, vem se opor a um outro uso até então dominante. (Aléong, 2001: 155)

Mais adiante (p. 173), o mesmo autor afirma:

> Quem quer que já tenha folheado uma obra de correção de erros sabe que a grande maioria dos erros não representam nenhuma dificuldade de compreensão. Seu erro é, simplesmente, tudo aquilo que, num dado momento, "não se diz". E nada impede que em outra época um erro se torne a regra.

8.1.5 Outra confusão: língua escrita = norma-padrão

A depreciação da fala como lugar do erro e da falta de gramática leva à situação descrita por Geraldi (1996: 62):

> [...] a escrita passou a ser usada como forma de "normatizar" a fala: para sujeitos letrados, o lugar utópico em que gostariam de estar quando falam (e que se imaginam ocupando quando falam) é "falar como se escreve" porque a escrita é que seria a "língua correta".

Nessa utopia, a escrita é tomada como um bloco homogêneo, invariável, sempre imaginada como a escrita literária ou de gêneros textuais mais monitorados e mais valorizados socioculturalmente ou, melhor, por um segmento privilegiado da sociedade.

Assim se explica por que as gramáticas normativas, em geral, não admitem outro modelo de língua a ser imitado senão o dos "grandes escritores". Os autores de compêndios gramaticais filiados à TGP consideram como "língua culta" apenas a língua *escrita* e, mais precisamente, a língua escrita pelos "bons autores", escolhidos segundo critérios de gosto estético totalmente subjetivos.

Esse modelo exclusivo é apresentado explicitamente por Cunha e Cintra (1985: xiv) no prefácio de sua obra:

> Trata-se de uma tentativa de descrição do português atual *na sua forma culta*, isto é, da língua como a têm utilizado os *escritores* portugueses, brasileiros e africanos do Romantismo para cá. [grifos meus]

Esse compêndio, portanto, teoricamente só poderia ser consultado por quem tivesse dúvidas no momento de *escrever um texto literário*, já que, segundo os próprios autores, não serão abordados nele fenômenos característicos de outras variedades escritas, como a jornalística ou a da produção científica, muito menos os fenômenos típicos da língua falada. No entanto, essa mesma obra traz um extenso capítulo sobre fonética e fonologia, ou seja, um exame das características *orais e acústicas* da língua, o que está em franco desacordo com o projeto traçado no prefácio do livro, que é o de estudar a língua dos *escritores*.

A maioria das obras desse gênero literário que é a gramática normativa se configuram de modo semelhante. Seus autores assumem a variedade escrita literária como a única digna de ser estudada, ensinada e praticada, e consideram isso tão natural que nem se dão ao trabalho de defini-la como seu objeto de estudo. Fica evidente que para eles só essa variedade escrita mais conservadora merece o rótulo de "língua portuguesa". O que vem estampado nesses compêndios vale, supostamente, para todas as línguas do grupo portugalego, em qualquer lugar do mundo, em qualquer momento histórico, em qualquer classe social, em qualquer faixa etária.

Os autores de gramáticas prescritivas, incluindo as mais recentes, não fazem a distinção básica entre *ortografia* e *fonética*, isto é, entre as regras da escrita oficial da língua e os fenômenos da língua oral. Isso fica ainda mais claro quando eles incluem a ortografia no mesmo capítulo em que apresentam a fonética.

Por isso é que a gente não se espanta quando os autores de uma gramática destinada ao uso escolar (Cipro e Infante, 1997: 35) explicam que a letra X representa o fonema /ʃ/ depois de um ditongo e dão como exemplo de palavras "com ditongo": AMEIXA, CAIXA, PEIXE, EIXO, FROUXO, TROUXA, BAIXO — sem mencionar a monotongação que se processou nessas palavras em todas as variedades do PB, incluindo as

VUP, resultando nas pronúncias [a'meʃa], ['kaʃa], ['peʃi], ['eʃu], ['froʃu], ['troʃa], ['baʃu]. O termo *ditongo* ('dois sons'), que se aplica a um fenômeno *fonético*, não cabe nesses exemplos, que retratam simplesmente a *convenção ortográfica* que ainda conserva, na *escrita*, as duas letras vogais antes do X. Trata-se, na verdade, de *dígrafos* ('duas letras'), termo mais adequado para descrever regras ortográficas.

Os mesmos autores alegam que na palavra QUAL existe um "ditongo crescente" (supostamente [ʊa]). Ora, se fosse para usar a terminologia tradicional, melhor seria falar de *tritongo*, pois o que de fato se ouve, na ampla maioria das variedades linguísticas do PB, é ['kʷaw], devido à total vocalização do que ainda se escreve L em sílaba travada e em final de palavra [▶148]. No entanto, como já vimos em [▶321], é preciso abandonar a noção de "tritongo", pois o que de fato existe nesses casos são *consoantes complexas*, fortemente velarizadas.

8.1.6 A ortografia não faz parte da língua

Seria fundamental que as professoras e os professores se conscientizassem de que saber ortografia não é saber a língua. São dois tipos diferentes de conhecimento, controlados, aliás, por partes diferentes do cérebro. Como vivemos numa cultura altamente grafocêntrica — centrada na escrita —, ignoramos o fato de que ¾ das línguas do mundo não têm forma escrita (só no Brasil são mais de 150), são chamadas línguas ágrafas, mesmo sendo tão complexas (e às vezes até mais complexas morfossintaticamente) do que as línguas mais difundidas mundialmente.

Fica óbvio, então, que a ortografia não faz parte da língua. Saber ortografia é como tocar piano, dançar balé, atirar com arco e flecha, manejar um programa de computador — são atividades que exigem treinamento, prática constante, memorização consciente. Saber a língua é outra coisa. Afinal, milhões de pessoas nascem, crescem, vivem e morrem sem jamais aprender a ler/escrever sendo, no entanto, conhecedoras perfeitas da *gramática* de sua língua, capazes de distinguir com clareza uma construção agramatical de uma gramatical.

É comum os linguistas se referirem aos idiomas humanos como línguas *naturais*. Podemos então, em contraposição a isso, afirmar que a ortografia é *artificial*, depende da vontade dos homens e, principalmente, dos homens que legislam sobre ela. Com muita frequência, ela fica sujeita aos gostos pessoais ou às interpretações dos fenômenos linguísticos por parte dos filólogos que ajudam a estabelecê-la.

8.1.7 Consequências da falácia

Uma das muitas consequências indesejadas da falácia clássica, na sua vertente da "superioridade" da escrita sobre a fala, é a crença, muito difundida sobretudo no

ambiente escolar, de que todo e qualquer texto escrito tem de ser sempre "formal" e "rebuscado" — escrever é usar ao máximo o que parecer mais "sofisticado" e menos "coloquial". A mentalidade que preside esse tipo de comportamento se traduz assim: "Quanto mais distante da fala diária, quanto menos habitual for determinada forma linguística, mais ela deve ser empregada na escrita". A tradição prescritivista do nosso ensino de língua tenta impedir que as pessoas utilizem sua eficiente intuição linguística na hora de produzir seus textos escritos.

Por exemplo, como persiste no nosso ensino a irracionalidade das regras de colocação pronominal [▸741] em que a *ênclise* — o pronome depois do verbo — é a "colocação básica" dos clíticos em português, embora de fato seja a próclise ao verbo principal a única regra em vigor na gramática do PB, muitas pessoas, em sua produção escrita, empregam única e exclusivamente a ênclise, mesmo em situações em que a própria TGP recomenda o uso da próclise:

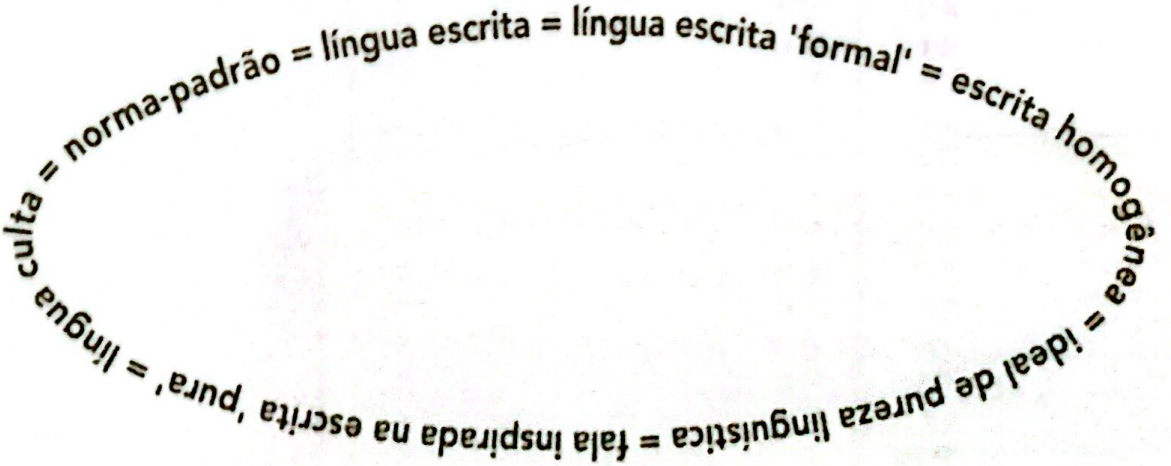
(10) E agora, pasmem: recebi um relatório ultrassecreto, que estava prestes a ser publicado no WikiLeaks, no qual insinua-se que, por trás de toda essa confusão [...] estaria... adivinhem quem? Sim, ele mesmo, o JUNIOR. (Marcio Alemão, *Carta Capital*, nº 626, 15/12/2010, p. 73)

(11) Encerrando o artigo, discutiu-se quem são as pessoas competentes para requerer a interceptação telefônica, onde viu-se que a Polícia Militar, na pessoa de seus oficiais, não está no rol do artigo 3º da lei de Interceptação Telefônica. (Alexsander Bernardes de Souza, "Lei de Interceptação Telefônica e a imparcializade do juiz" <http://www.lfg.com.br/public_html/article.php?story=20100310092147374&mode=print>, acesso em 17/12/2010)

(12) SITE – ficou fora do ar por dois dias, em função de que era compartilhado com o provedor do site do presidente anterior, sr Marcelo. Informou que o srº Adrien que era o responsável pela alimentação do site, não encontra-se mais trabalhando, pois seu contrato encerrou-se em 30/04/10. (<http://www.cogemaspr.org.br/Noticias/noticia054.aspx>, acesso em 17/12/2010)

No capítulo 21, vamos nos ocupar mais demoradamente desses casos de hipercorreção.

A "falácia clássica", portanto, produziu ao longo dos séculos um círculo de sinonímias equivocadas, que podemos representar da seguinte maneira:



Embora seja fácil demonstrar a falácia clássica, ela só é vista como falaciosa por uma reduzida minoria de pessoas. Apesar dos esforços dos linguistas e de outros estudio-

sos contemporâneos em provar que a fala e a escrita têm muito mais semelhanças do que diferenças entre si, a língua escrita permanecerá sempre, no imaginário coletivo, no senso comum, como algo superior, mais sublime e mais digno de veneração do que a língua falada. O discurso social sobre a escrita, sobre o que está registrado "preto no branco", é há milênios um discurso de supervalorização do escrito e de superdepreciação do falado. Em sociedades entranhadamente grafocêntricas, ou mais precisamente, organizadas em torno de um poder definido e controlado pelos que sabem/podem ler e escrever (sempre uma minoria), o que se reserva para a oralidade é, quando muito, o terreno do pitoresco, do folclórico, do anedótico. Aqui também vale a discussão já feita no início do capítulo 3, sobre língua materna (falada) e língua paterna (escrita).

8.2 O ALFABETO

8.2.1 Um pouco de história

O grafocentrismo que caracteriza a cultura ocidental — e as sociedades que, por obra do colonialismo e do imperialismo, tiveram de se enquadrar nela — obscurece para a grande maioria das pessoas a história complexa da instituição dos sistemas de escrita e das normas ortográficas vigentes em cada sociedade. Também obscurece as vicissitudes políticas e ideológicas que dirigem essa instituição.

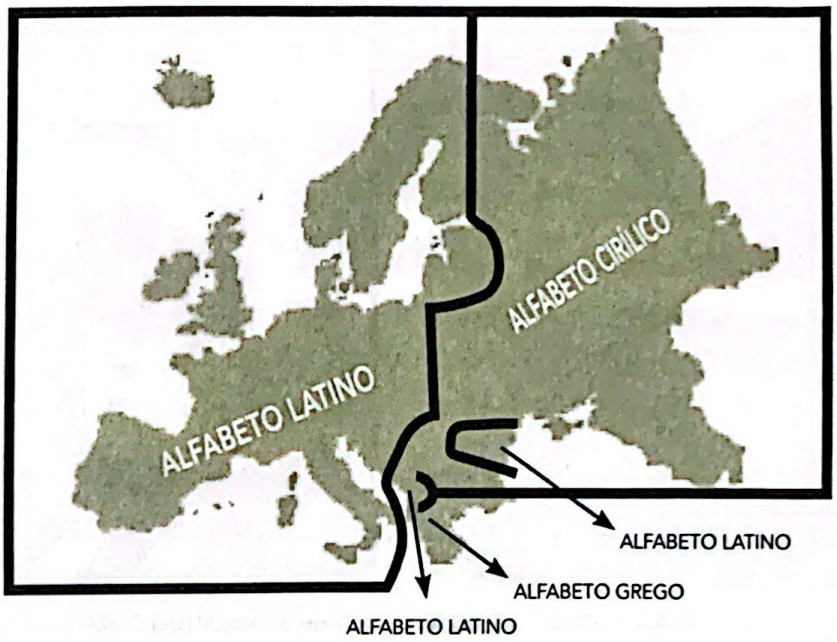
Por exemplo, a distribuição geográfica dos dois alfabetos mais empregados na Europa — o alfabeto latino e o alfabeto cirílico — reflete importantes fatos históricos que remontam a muitos séculos atrás.

ALFABETO LATINO	
A a	N n
B b	O o
C c	P p
D d	Q q
E e	R r
F f	S s
G g	T t
H h	U u
I i	V v
J j	W w
K k	X x
L l	Y y
M m	Z z

ALFABETO CIRÍLICO			
А а	А а	*Р р*	Р р
Б б	Б б	*С с*	С с
В в	В в	*Т т*	Т т
Г г	Г г	*У у*	У у
Д д	Д д	*Ф ф*	Ф ф
Е е	Е е	*Х х*	Х х
Ж ж	Ж ж	*Ц ц*	Ц ц
З з	З з	*Ч ч*	Ч ч
И и	И и	*Ш ш*	Ш ш
Й й	Й й	*Щ щ*	Щ щ
К к	К к	*Ъ ъ*	Ъ ъ
Л л	Л л	*Ы ы*	Ы ы
М м	М м	*Ь ь*	Ь ь
Н н	Н н	*Э э*	Э э
О о	О о	*Ю ю*	Ю ю
П п	П п	*Я я*	Я я

O mapa ao lado deixa bem claro que existe uma nítida separação entre Europa ocidental e Europa oriental no que diz respeito ao uso dos alfabetos: no ocidente, o latino, e

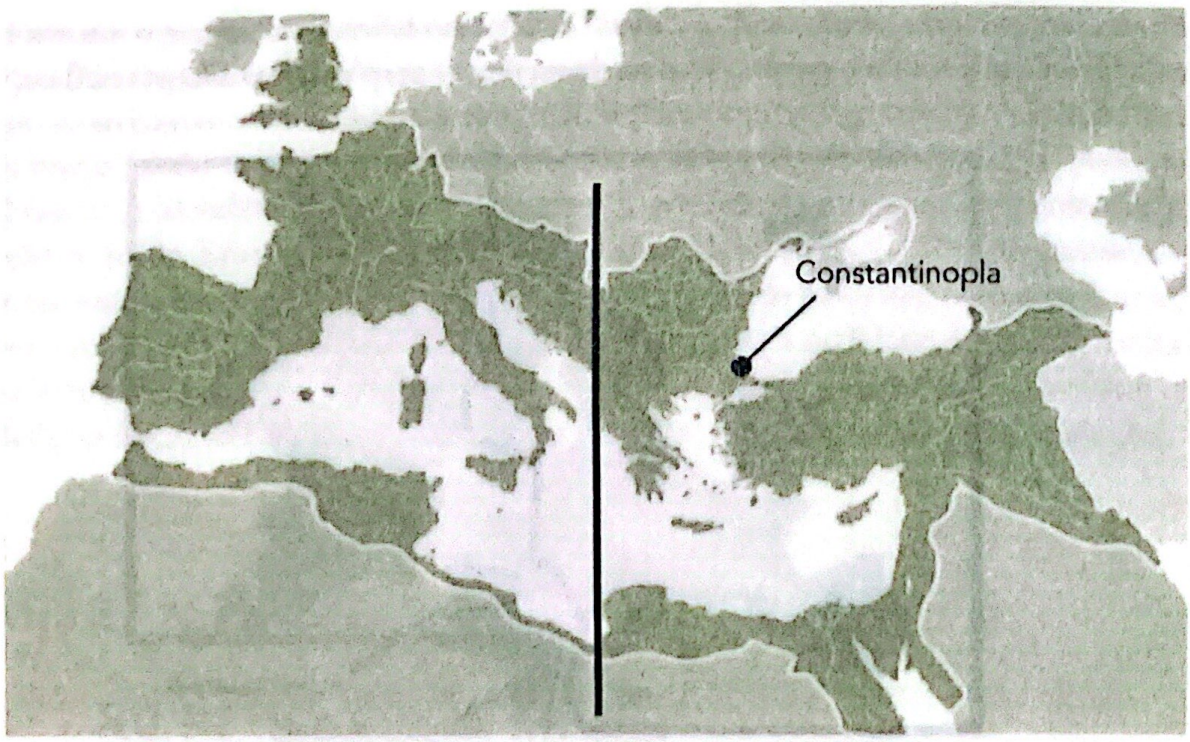
no oriente — com exceção de duas "ilhas" de alfabeto latino (o albanês e o romeno) e da Grécia, que tem seu próprio sistema de escrita — prepondera o alfabeto cirílico.



O que terá provocado essa divisão? A resposta é muito simples: a religião. No ano de 313, o imperador romano Constantino (272-337) publicou o célebre Edito de Milão, que assegurava tolerância religiosa aos cristãos do império. Até então, os cristãos tinham sido alvo de perseguições e massacres. Constantino também será responsável pela transformação da cidade de Bizâncio, na extremidade oeste da parte europeia do império, numa residência oficial do imperador com o nome de Constantinopla. Essa cidade se tornaria em breve a capital do Império Romano do Oriente e permaneceria com esse *status* por mais de mil anos, até ser conquistada pelos turcos, que lá estão até hoje e a chamam de Istambul.

Pouco mais tarde, o imperador Teodósio (347-395) transforma a tolerância religiosa em oficialização do cristianismo como única religião permitida no Império. Teodósio vai ser o último governante do Império Romano unificado — após sua morte, aquele vasto território será dividido em duas porções: o Império Romano do Ocidente, com sede em Milão, e o Império Romano do Oriente, com sede em Constantinopla (o império oriental se tornará conhecido mais tarde como Império Bizantino, por causa do nome mais antigo da capital, Bizâncio).

Essa divisão tem importantes consequências culturais, sobretudo linguísticas e religiosas. A parte oriental do Império era dominada pela língua e pela cultura gregas depois das conquistas de Alexandre Magno no século IV a.C. A Grécia foi subjugada pelos romanos a partir de 168 a.C., a Ásia Menor, o Oriente Médio e o Egito, no século I a.C. Mesmo submetida ao domínio político romano, toda essa área continuou preservando a língua grega e a cultura helenística.



[Divisão do Império Romano em Ocidente e Oriente (séc. IV d.C.).]

Uma evidência desse predomínio da língua grega como idioma de integração da parte oriental do Império Romano é o fato de todos os textos incluídos no Novo Testamento terem sido escritos em grego, muito embora seus autores fossem judeus vivendo sob o governo de Roma. No norte da África, a linha divisória também refletia essa divisão linguística: enquanto no Egito (principalmente em Alexandria) floresceu toda uma geração de teólogos e historiadores que se expressou em grego, do outro lado da linha, surgiram grandes autores cristãos que se expressaram unicamente em latim (sendo o mais importante deles santo Agostinho, nascido na cidade de Hipona, que hoje pertence à Argélia e se chama Anaba).

Com a divisão do Império, cada uma das duas partes foi desenvolvendo suas próprias tradições culturais e religiosas. O Império Romano do Ocidente desmoronou a partir do século V, quando foi conquistado por vários grupos étnicos diferentes, principalmente germânicos (visigodos, ostrogodos, lombardos, alamanos, burgúndios, alanos, vândalos, francos etc., no continente, e, nas Ilhas Britânicas, anglos, saxões, jutos, viquingues, daneses etc.), que estabeleceram reinos independentes.

A historiografia moderna evita se referir a esse movimento como "invasões bárbaras", pois o que de fato parece ter ocorrido foi uma grande migração de povos que viviam além das fronteiras do Império. De fato, parece que, ao contrário do que sempre se alegou, essas migrações foram uma consequência e não a causa do declínio do Império do Ocidente, que já estaria em decadência nesse período.

Apesar do desaparecimento do Império do Ocidente, o cristianismo foi adotado pelos povos germânicos, junto com a língua latina. Roma, a antiga capital impe-

rial, se tornou a capital "espiritual" do Ocidente. O bispo de Roma, chamado de papa (termo infantil para *pai*), assumiu o lugar dos imperadores e passou a reinar "espiritualmente" sobre toda a população cristianizada do Ocidente. O papa também exercia grande poder político: qualquer monarca só poderia ser considerado digno de governar se recebesse a bênção papal e o reconhecimento de seu Estado pela Santa Sé. O papa também adotou para si os símbolos do poder imperial como o trono, a mitra, o manto etc. E até hoje o papa é, do ponto de vista do direito internacional, um monarca absolutista, chefe de Estado do Vaticano, que só deixa seu posto ao morrer.

No Império do Oriente, que sobreviverá até 1452, ao ser conquistado pelos turcos muçulmanos, o imperador preservou o poder político enquanto o poder "espiritual" era exercido pelo arcebispo de Constantinopla, que passará a ser chamado de patriarca e, mais tarde, Patriarca Ecumênico de Constantinopla.

Com o passar do tempo, as duas igrejas — a romana e a bizantina — entrarão em conflito quanto à divisão de seus poderes, às suas áreas de influência e quanto aos dogmas que cada uma delas vai defender. Essa divisão também se reflete em termos culturais e linguísticos: a igreja romana preserva o latim como seu idioma oficial (até hoje), enquanto a igreja bizantina conservou o grego. Os desentendimentos entre as duas igrejas culminarão no ano de 1054, quando ocorrerá o chamado Grande Cisma, em que o cristianismo se dividirá entre a Igreja Católica Romana e a Igreja Católica Ortodoxa.

Já antes dessa grande separação, no entanto, cada uma das capitais religiosas difundia sua língua e sua cultura litúrgica nos territórios sob sua jurisdição. No Ocidente, missionários católicos romanos foram enviados às Ilhas Britânicas e à Irlanda (são Patrício, 387-493), às terras habitadas pelos germânicos (são Bonifácio, 672-751): essas regiões se tornaram católicas, mas o latim não se impôs como língua das populações. Entre os séculos IX e X, foi a vez de alguns povos eslavos (tchecos, boêmios, morávios, poloneses, eslovenos, croatas) caírem sob a influência religiosa de Roma. No século XIII, os últimos pagãos da Europa — lituanos, letões e estonianos — foram convertidos pela força das armas dos cruzados vindos da atual Alemanha. No ano 1000, o rei da Hungria, Estêvão I (santo Estêvão), se converte ao catolicismo romano. Assim, em todos esses territórios também se impôs o alfabeto latino.

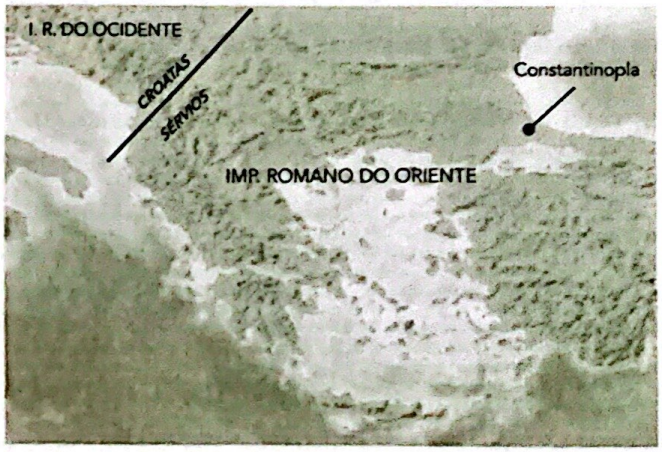
Do lado oriental, a igreja de Constantinopla também cuidava de sua expansão. Sob sua jurisdição estavam diversos povos eslavos que tinham iniciado uma migração desde suas terras ancestrais (provavelmente a atual Ucrânia) rumo ao leste e ao sul. Para catequizar esses povos, foi criado o alfabeto cirílico, assim chamado em homenagem a são Cirilo (826-869), um religioso bizantino que com seu irmão Metódio (815-885) formará a dupla dos chamados "Apóstolos dos eslavos". A criação do alfabeto, no entanto, não pode ser inteiramente atribuída a ele. Esse alfabeto se formou tendo como inspiração as letras do alfabeto glagolítico (de origem controversa, usa-

do por alguns eslavos) e do alfabeto grego. Por isso, os povos eslavos cristianizados pelos bizantinos se filiam até hoje à Igreja Ortodoxa e empregam o alfabeto cirílico: russos, bielo-russos, ucranianos, búlgaros, sérvios, macedônios, montenegrinos. Em 988, o príncipe Vladimir se converteu ao cristianismo bizantino e impôs a religião a todo o povo russo, sob seu domínio. Com o surgimento da União Soviética no início dos século XX (1917), o alfabeto cirílico, em sua versão russa, foi imposto a todos os povos que compunham a gigantesca federação (cazaques, quirguizes, tadjiques, uzbeques, moldavos, turcomanos etc.).

Na Albânia, país que sempre enfrentou invasões estrangeiras das mais diferentes procedências, a língua podia ser escrita em alfabeto latino (pelos católicos do norte), em alfabeto cirílico (pelos ortodoxos do sul) e em alfabeto árabe (pelos muitos muçulmanos do país, convertidos ao islamismo durante a dominação otomana). Em 1908, foi adotado o alfabeto latino para todo o país.

Na Romênia, país circundado por povos de línguas eslavas, o alfabeto cirílico foi empregado até o final do século XIX. Nesse período, os linguistas romenos estabeleceram definitivamente a origem latina de sua língua (a única língua românica falada na Europa oriental) e, com isso, o alfabeto latino se impôs, tendo como base os usos das letras feitas em italiano. A maioria da população romena, no entanto, é cristã ortodoxa.

Esses desdobramentos históricos explicam uma situação bastante curiosa. A língua servo-croata (eslava) sempre foi considerada pelos linguistas como uma só. No entanto, sérvios e croatas sempre se diferenciaram por dois aspectos importantes: os sérvios são cristãos ortodoxos e escrevem sua língua com o alfabeto cirílico; os croatas são católicos romanos e, portanto, escrevem com alfabeto latino. Essas diferenças se devem a um fato histórico muito simples: quando o Império Romano foi dividido em ocidental e oriental, a linha divisória passou exatamente no meio do território que, mais tarde, seria ocupado por uma população eslava. Com isso, a porção norte daquela área foi cristianizada por missionários romanos, enquanto a porção sul foi cristianizada por missionários bizantinos. A secular rivalidade que sempre opôs croatas a sérvios (até hoje) é a continuação histórica das disputas entre Roma e Constantinopla pela conquista de novos rebanhos para suas igrejas.



Outro sistema de escrita que se difundiu amplamente por motivos religiosos foi o alfabeto árabe. A partir do século VII, com a expansão do islamismo, todo o norte da África foi conquistado pelos árabes (e também a Península Ibérica [▸217]). A língua árabe, sua forma

de escrita e a religião islâmica se impuseram naquela região, expulsando para sempre o grego do Egito e o latim das outras províncias norte-africanas. Outros povos, falantes de outras línguas sem nenhum parentesco com o árabe, assumiram a religião islâmica e, com ela, também adotaram o alfabeto árabe: turcos, persas, afegãos, paquistaneses, uigures (na China), curdos (Turquia e Iraque), jawis (Tailândia e Brunei), maldivos etc. Na Turquia, o alfabeto árabe foi empregado até 1928, quando foi substituído pelo latino, num processo de ocidentalização do país com vistas a aproximá-lo mais da Europa. E o país com a maior população islâmica do mundo, a Indonésia (235 milhões), emprega o alfabeto latino, por ter sido uma colônia holandesa.

8.2.2 O alfabeto latino da primeira à última letra

O alfabeto latino é hoje o mais empregado em todo o mundo. Além da Europa Ocidental, os países surgidos dos impérios coloniais europeus também empregam esse alfabeto: todo o continente americano, a maior parte da África negra, alguns países asiáticos (Indonésia, ex-colônia holandesa; Malásia, ex-colônia britânica; Filipinas, ex-colônia espanhola; Timor Leste, ex-colônia portuguesa; Vietnã, ex-colônia francesa), toda a Oceania. Além disso, em línguas como o chinês e o japonês se desenvolveu alguma forma de transcrição que emprega o alfabeto latino: o *pinyin* para o chinês e o *rōmaji* para o japonês.

O alfabeto originalmente usado pelos romanos era o seguinte:

A B C D E F Z H I K L M N O P Q R S T V X

Esse alfabeto se inspirava no alfabeto etrusco, por sua vez inspirado no alfabeto grego ocidental. O grego, por seu turno, foi construído com base no sistema de escrita fenício, tendo herdado deste inclusive o nome de algumas letras:

LETRA GREGA	NOME GREGO	NOME FENÍCIO
Α α	alfa	'alef
Β β	beta	beth
Γ γ	gama	gimel
Δ δ	delta	daleth
Θ θ	teta	teth
Κ κ	kapa	kaf
Λ λ	lambda	lamedh
Τ τ	tau	taw

A própria palavra *alfabeto* provém diretamente do grego, do nome das duas primeiras letras. Outro nome que empregamos também é *abecedário* ou *abecê*, do nome das três primeiras letras do alfabeto latino.

A letra C, no latim arcaico, representava tanto o som [k] quanto o som [g]. Isso se deve, provavelmente, ao fato de não existir no etrusco o som [g]. Para resolver essa ambiguidade, os romanos criaram (no séc. III a.C.) a letra G, que é, de fato, um C com uma pequena barra horizontal. Como não tinham necessidade da letra Z, que representava um som inexistente em latim, os romanos excluíram essa letra e puseram em seu lugar, na ordem alfabética, o G.

A letra K era de uso muito restrito, como na palavra *kalendae* (primeiro dia de cada mês), de onde provém o nosso *calendário*. A palavra *calendas* é usada na expressão "ficar para as calendas gregas", isto é, para um dia que jamais virá, pois os gregos não tinham calendas, só os romanos.

Depois da conquista da Grécia por Roma, os romanos adotaram as letras gregas Y e Z (de fato readotada) para transcrever empréstimos do grego. Isso explica por que elas são as últimas do alfabeto.

A letra Q era chamada *qoppa* pelos gregos, que usavam ela para representar o som [k] diante das vogais O [õ], Y [y] e Ω [ō], isto é, vogais posteriores [▶292] (o símbolo [y] corresponde ao U do francês e ao Ü do alemão). No entanto, como já existia a letra K, logo o *qoppa* foi abandonado pelos gregos. Os romanos se serviram dela para representar a consoante complexa [k^w], que é um único fonema, distinto da sequência [kʊ]. Assim, *quidam* ('algum') ['k^widam], se diferencia de *cuius* ('cujo') ['kuɪus].

Desse modo, no século I a.C., no chamado período clássico da língua latina, o alfabeto tinha suas 23 letras mais difundidas:

A B C D E F G H I K L M N O P Q R S T V X Y Z

Ao contrário dos gregos, que davam um nome próprio a cada letra, muitos deles, como vimos, oriundos do fenício, os romanos chamavam as letras de seu alfabeto de modo diferente:

Letra	Nome
A	ā
B	bē
C	cē [ke:]
D	dē
E	ē
F	ef
G	gē
H	hā
I	ī
K	kā
L	el
M	em

N	en
O	ō
P	pē
Q	qū [k^{w}u]
R	er
S	es
T	tē
V	ū
X	ex [eks]
Y	ī Graeca
Z	zēta

Os romanos, portanto, tinham uma boa consciência do caráter soante de L, M, N, R e contínuo de F e S, de modo que para designar essas letras usavam uma vogal [e] *antes* do som que elas representavam: [el], [em], [en], [er] etc. As vogais também eram chamadas pelo som: [a:], [e:], [i:], [o:], [u:] (o símbolo [:] indica que eram vogais longas). Para as consoantes, que precisavam de apoio para serem emitidas, acrescentavam uma vogal [e:] *depois* do som: [be:], [ke:], [de:] etc. A letra X era chamada [eks] e representava, de fato, a junção K+S ou G+S. O Y era chamado de "i grego" porque representava um som inexistente em latim, som que, com o passar do tempo, se transformou num simples [i]. O Z era chamado por seu nome grego: *zeta*.

Os romanos atribuíam às letras o gênero feminino, pois ao se referirem a elas subentendiam a palavra *littera* ('letra'). Essa tradição é seguida até hoje pelo italiano e pelo espanhol, ao passo que o português e o francês optaram pelo masculino.

As pessoas habituadas aos alfabetos modernos derivados do latino decerto estranharão algumas ausências na série de letras do abecedário mais antigo usado pelos romanos. Vamos ver isso.

A letra V dos romanos representava a vogal [u] ou a semiconsoante [ʊ]. Seu desenho foi inspirado no Y grego. No latim clássico não existia a consoante [v]. Com o passar do tempo, algumas línguas românicas desenvolveram essa consoante, resultante da consonantização total da semiconsoante [ʊ]: *uacca* > *vaca*. Apesar do surgimento do novo som, a letra V demoraria muito para representá-lo. Na escrita medieval tardia, e mesmo na renascentista, se usava V no início das palavras e U no meio ou no final, sem consideração pelo som que elas representavam. Isso ocorre, por exemplo, na edição original de *Os Lusíadas* (1572):

> Cessem do sabio Grego, & do Troyano,
> As nauegações grandes que fizerão;
> Callese de Alexandre, & de Trajano,
> A fama das victorias que tiuerão,
> Que eu canto o peyto illustre Lusitano,
> A quem Neptuno, & Marte obedeçerão.

Cesse tudo o que a Musa antigua canta,
Que outro valor mais alto se alevanta. (I, 3)

Assim como a letra J, que vamos ver adiante, o V é chamado de letra ramista em referência ao filósofo francês Pierre de la Ramée (em latim, Petrus Ramus, 1515-1572) que foi o primeiro a empregá-las como distintas do I e do U.

O som que se desenvolveu a partir do [ʊ] latino foi o [v], que representamos hoje com a letra V. Esse som está presente no italiano do centro-norte, no romanche (Suíça), no francês, no provençal, no romeno, no português europeu e nas línguas dele derivadas, como o PB. Em outras línguas, porém, o som que se firmou foi [b] ou [β]: sardo, italiano do sul, espanhol, catalão, galego, gascão. Assim, o português é a única língua românica ibérica que apresenta a distinção [v]/[b], embora nos dialetos do norte de Portugal, mais próximos da Galiza, essa distinção não ocorra. A ortografia do espanhol, no entanto, conserva tanto o V quanto o B etimológicos, apesar dos sons idênticos, o que cria dificuldades para muitas pessoas falantes da língua no momento de escrever. A neutralização da diferença entre [b] e [v] a favor de [b] ou [β] é chamada de *betacismo*. Alguns autores postulam que o betacismo ibérico, do qual só o português do centro-sul escapa, se deve ao contato linguístico com o basco, língua em que não existe a consoante [v].

Outra letra que tem sua origem remota no Y é o W (dábliu). Como o próprio nome indica, em inglês, trata-se de um "duplo U" (*double U*). Essa letra não é empregada nas línguas românicas, a não ser para grafar palavras estrangeiras. Com a evolução do V [u]/[ʊ] latino em [v], essa letra deixou de servir para representar o som labiovelar aproximante das línguas germânicas cujo símbolo fonético é [ɥ]. Esse som era muito velarizado no germânico antigo e, por isso, as palavras de origem germânica que o apresentavam e que se introduziram nas línguas românicas foram grafadas com G ou GU:

GERMÂNICO	LATIM MEDIEVAL	PORTUGUÊS
wardon	guardare [gʷar'dare]	guardar
warnjan	guarnire [gʷar'nire]	guarnir
werra	guerra ['gʷeřa]	guerra
widan	guidare [gʷi'dare]	guiar
wisa	guisa ['gʷisa]	guisa

E os nomes próprios germânicos que continham [ɥ] passaram às línguas românicas escritos com G:

Wit > Guido
Wilhelm > Guilherme
Weigmahrs > Guimara (> Guimarães)
Wigmar > Guiomar
Waldhar > Guálter

Para grafar o som [ɯ], os escribas de línguas germânicas passaram a usar VV ou UU, que logo se transformariam em W. O alemão moderno não tem mais o som [ɯ], que se transformou em [v] e que continua sendo grafado com w. Em inglês o som velar permaneceu, também escrito W.

A letra I do alfabeto romano representava tanto a vogal [i] quanto a semivogal [ɪ]. Na formação das línguas românicas, o [ɪ] se transformou numa consoante palatal [ʤ] que recebeu a grafia J que, de fato, é um I com uma pequena curva para a esquerda. Na forma minúscula, o *j* conservou o "pingo" do *i*. Também foi Pierre de la Ramée o primeiro a distinguir o I do J. Essa consoante manteve seu som africado [ʤ] em italiano, onde é grafado GE e GI como em *gente* ['ʤente]. Em português e francês, o J, surgido da semiconsoante latina, assumiu o som [ʒ], enquanto GE e GI eram pronunciados como [ʤe] e [ʤi]. Assim, no português arcaico, não havia dúvida quanto à grafia de uma palavra com J ou GE/GI, porque os sons que essas letras e sílabas representavam eram distintos. Com o passar do tempo, porém, GE/GI perderam seu traço africado e se fundiram ao [ʒ] escrito com J, tanto em português quanto em francês. Daí os problemas que temos hoje no momento de escrever alguma palavra pouco conhecida com o som [ʒ] seguido de [e] ou [i].

As letras C e G do latim clássico representavam sempre os sons do nosso C em *cacareco* e do nosso G em *gargalo*. Por que é que hoje pronunciamos a palavra *certo* como ['sɛhtu] e não como ['kɛhtu], e a palavra *giro* como ['ʒiru] e não como ['giru]? Esse fenômeno não é exclusivo do PB nem do português europeu: ele se verifica em todas as línguas da família românica e também em outras que adotaram o alfabeto latino, como o inglês e o alemão (só no caso do C). Repetindo o que já dissemos: se você prestar atenção ao modo como fala, perceberá que as sílabas [kɛ], [ke], [ki] e [gɛ], [ge], [gi] não são pronunciadas da mesma forma como pronunciamos [ka], [kɔ], [ko], [ku] e [ga], [gɔ], [go] e [gu]. Como as vogais daquelas primeiras sílabas são produzidas perto do palato, as consoantes [k] e [g], que são pronunciadas na região velar (onde está o véu palatino também chamado de palato mole), são "puxadas" para adiante, adquirindo um som palatalizado. Assim, o trajeto percorrido pelo som [k] seguido de [ɛ], [e] e [i] e do [g] na mesma situação foi o seguinte:

[ki] > [kʲi] > [ʦi] > [si]
[gi] > [gʲi] > [ʤi] > [ʤi] > [ʒi]

Essas mesmas etapas ocorreram na história do francês. No castelhano e no galego, a etapa [ʦ] foi seguida pela mudança em [ɵ], a consoante interdental surda (como o TH do inglês em *think*). Nas outras línguas do grupo hispânico (argentino, chileno, colombiano, mexicano etc.), essa interdental não existe, de modo que CE e CI são pronunciados como no PB. Por sua vez, o antigo som [ʒ] que existia no castelhano se transformou na aspirada [x], que se escreve com J ou com GE, GI em todas as línguas do grupo. No toscano (base do italiano padrão), a etapa [ʦi] foi seguida de uma nova

palatalização que resultou em [ʧ]. Já no caso do C, o toscano se deteve na fase [ʤ]. O apego à herança latina é que explica a permanência das grafias CE, CI, GE, GI que poderiam ter sido substituídas por SE, SI, JE, JI, o que decerto facilitaria bastante a nossa vida.

8.2.3 O alfabeto latino, seus diacríticos e seus dígrafos

O emprego do alfabeto latino para grafar línguas diferentes do latim — incluindo as próprias línguas românicas, derivadas dele — e, principalmente, línguas de outros grupos e famílias linguísticas, fez surgir uma série de problemas que, em cada lugar, receberam soluções distintas, de acordo com a história e as características de cada sociedade. O exemplo da invenção do W, que vimos acima, ilustra bem esse fato.

Outro exemplo é o da letra J. No alemão, no holandês e nas línguas escandinavas (dinamarquês, sueco, norueguês e islandês), nas línguas eslavas que adotaram o alfabeto latino (polonês, tcheco, eslovaco, esloveno, croata), no húngaro, no finlandês, no albanês, no lituano, no letão e no estoniano, a letra J representa a semivogal [ɪ], que escrevemos com a mesma letra I que usamos para representar a vogal [i]. Por isso, a palavra norueguesa *fjord* foi aportuguesada em *fiorde*, o nome do psicanalista suíço *Jung* se pronuncia [nuŋk] e a capital da Eslovênia, Ljubljana, é escrita Liubliana em português. É provável que todas essas línguas tenham se inspirado no uso alemão do J para representar a semivogal [ɪ]. No entanto, várias delas têm o som [ʒ], que se representa em francês e português como J. Assim, para grafar o som [ʒ] surgiu em cada língua uma opção diferente, por exemplo:

- albanês → zh
- húngaro → zs
- lituano → ž
- polonês → ż
- tcheco → ž

O [ʒ] é um dos muitos sons palatais que não existiam em latim. Outro som palatal muito comum é o [ʃ], como em *xixi*, que a língua de Cícero desconhecia. Para representá-lo com os recursos do alfabeto latino surgiram as seguintes grafias, entre outras:

- francês → ch
- italiano → sce, sci
- polonês → sz, ś
- português → ch, x
- catalão e galego → x
- romeno → ș
- tcheco → š

Também a nasal palatal [ɲ], inexistente em latim, recebeu diversas grafias:

- basco → in
- albanês → nj

- malaio → ny
- tcheco → ň
- francês, italiano → gn
- galego, espanhol → ñ
- occitano, português → nh
- norueguês → nn

Os pequenos sinais que são acrescentados acima ou abaixo das letras convencionais para que elas representem sons diferentes dos das letras simples são chamados diacríticos. Algumas línguas recorrem a muitos diacríticos, como as eslavas que adotaram o alfabeto latino, o húngaro e o turco, línguas não indo-europeias:

- tcheco → ě č ď ň ř ů ť ý š
- húngaro → à ě í ó ö ü ű
- turco → ç ğ ş ö ü â î û

Outras línguas, como o inglês, não usam absolutamente nenhum diacrítico. Por isso, e por causa da imposição do inglês como língua internacional de comunicação devido ao domínio dos Estados Unidos sobre a política e a economia mundiais, é que somos obrigados a não usar diacríticos, por exemplo, nos endereços e senhas da internet.

Os diacríticos que usamos para escrever o PB são:

- o acento grave: à
- o acento agudo: á é í ó ú
- o acento circunflexo: â ê ô
- o til: ã õ
- a cedilha: ç

O trema, que era empregado apenas sobre o u (ü) deixou de ser usado depois da implementação do novo Acordo Ortográfico em 2009. (Mas é sempre bom lembrar que o trema não era um acento!)

Outra solução para suprir a necessidade de representação gráfica de sons inexistentes em latim são os dígrafos, isto é, duas letras para representar um único som. Nas línguas que adotaram o alfabeto latino os dígrafos são muito frequentes:

- basco: dd, ll, rr, tt, tx, tz
- húngaro: cs, dz, gy, ly, ny, sz, ty, zs
- polonês: ch, cz, dz, dź, rz, sź

Existem línguas que empregam, além de dígrafos, também trígrafos e quadrígrafos. É o caso do alemão (sch, tsch), do francês (eau, eaux, ain, aim, aou, aoû, aon, ein, ill), do húngaro (ccs, dzs, ggy), do holandês (aai, eeuw, ieuw), do inglês (igh, ssi, ough) entre outras línguas. No PB também temos alguns trígrafos (ver a seguir).

Os dígrafos que usamos para escrever o PB são:

DÍGRAFO	SOM	EXEMPLOS
1. AI	[a][1]	baixa - caixa
2. AM, AN	[ã]	amplo - campo - andar - cantar
3. CH	[ʃ]	chapéu - chuva - rachar - fecho
4. EI	[e][2]	beira - esteira - eira - queixo - queijo
5. EM, EN	[ẽ]	embora - lembrar - entrar - parente
6. GU, GO	[gʷ] ou [g]	guarda - mágoa - guerra - guia
7. HA, HÁ	[a]	haver - harmonia - hábito - hálito
8. HE	[e] ou [ɛ]	hera - herança
9. HÉ	[ɛ]	hélice - hélio
10. HI, HÍ	[i]	hibernar - hínico
11. HO	[o] ou [ɔ]	hotel - hospital - hora - horta
12. HÓ	[ɔ]	hóspede - hóstia
13. HU, HÚ	[u]	humilde - húmus
14. IG, ÍG	[i]	amigdalite - amígdala
15. IM, IN	[ĩ]	impor - limpar - jardim - índio - cinco
16. LH	[ʎ]	alho - velho - ilha - olho - hulha
17. NH	[ɲ]	manha - venha - tinha - sonha - unha
18. OM, ON	[õ]	ombro - sombra - batom - onde - íon
19. OU	[o]	ouro - estouro - cantou - chegou - vou
20. QU, CO	[kʷ] ou [k]	quadro - coalhar - quente - quinta
21. RR	[h], [x], [ř]	carro - barro - arroz - irrigar - birra
22. SC	[s]	nascer - crescer - ascensão - desci
23. SÇ	[s]	nasça - cresço - desço - floresça
24. SS	[s]	asso - isso - osso - russo - passa
25. UM, UN	[ũ]	umbral - alumbrar - algum - ungir
26. XC	[s]	exceção - exceder - excitar
27. XS	[s]	exsudar - exsurgir - exsucção

[(1) Diante da palatal surda [ʃ].]
[(2) Diante de [ʒ], [ʃ] e [ɾ].]

Além dos dígrafos, temos um caso particular de monógrafo, isto é, uma letra para representar dois sons: o x que pode representar [ks] (ou, melhor, [kˢs]) em palavras como *fixo*, *tóxico*, *axial* etc.

Por fim, temos uma letra — o H inicial — que não representa som nenhum e é mantida na ortografia por mera subserviência etimológica. O mesmo ocorre em francês e espanhol, mas não em italiano e romeno:

- latim: homine- → port.: homem ['õmẽi]
  esp.: hombre ['ombre]
  fran.: homme ['ɔm]
- latim: homo → ital.: uomo ['uomo]
  rom.: om ['om]

A manutenção do h em nossa ortografia é responsável pela existência dos seguintes trígrafos na escrita do PB:

TRÍGRAFO		SOM	EXEMPLOS
1. 2.	**ham** **han**	[ã]	hambúrger hangar - hanseníase
3.	**hen**	[ẽ]	hendecassílabo
4. 5.	**hin** **hín**	[ĩ]	hindu - hindustâni híndi
6. 7.	**hom** **hon**	[õ]	hombridade honra - hondurenho
8. 9.	**hum** **hún**	[ũ]	Humberto húngaro

8.3 A ORTOGRAFIA

8.3.1 A imprensa e a Reforma

Durante mais de mil anos, as línguas europeias não tiveram uma ortografia unificada, normatizada, instituída. Como durante esse longo período o latim foi a única língua estudada pelos intelectuais e empregada para a escrita de textos acadêmicos, científicos, religiosos etc., as línguas maternas das diferentes populações, quando eram escritas, tinham sua grafia inspirada no alfabeto latino, sem, contudo, nenhum rigor e, ao contrário, muito ao gosto do escriba.

Dois fatos históricos importantíssimos determinaram, em boa medida, a regularização da grafia dessas línguas: a invenção da imprensa e, em parte graças a ela, a Reforma protestante.

Por volta de 1439, o alemão Johannes Gensfleisch zur Laden zum Gutenberg empregou pela primeira vez os tipos móveis para imprimir livros, sendo o primei-

ríssimo deles a famosa Bíblia de Gutenberg. A invenção da imprensa é tida pelos historiadores como o fato mais importante da história moderna. Ela foi essencial para a eclosão do Renascimento, da Reforma protestante, da revolução científica, da economia baseada no conhecimento e da difusão do letramento de massa. E também, é claro, para a renovação da falaciosa superioridade da língua escrita.

Para imprimir se tornou necessário regularizar a grafia das línguas, dar a elas uma forma *precisa*, *direita*, ou seja uma ortografia (do grego *orthos*, 'reto, direito, justo', e *graphein*, 'escrever'). Não é por coincidência, então, que precisamente a partir desse período e, mais ainda, no Renascimento, as línguas europeias vão ser submetidas aos processos de normatização, de gramatização (produção das primeiras gramáticas) e de ortografização.

A Reforma protestante, que teve seu início simbólico em 1517 com os primeiros *protestos* de Martinho Lutero contra a devassidão do catolicismo romano, também contribuiu muito para os três processos listados acima. Os protestantes reivindicavam o direito de ler a Bíblia em suas próprias línguas maternas, sem a intermediação e a interpretação rígida imposta pela Igreja católica. Por isso, durante muitos anos, traduções da Bíblia circulavam clandestinamente e eram objeto de contrabando — muitos de seus tradutores foram presos, torturados e mortos, como o inglês William Tyndale, queimado na fogueira em 1535. Esse terrorismo católico, no entanto, não impediu a ampla difusão das ideias reformistas, que se impuseram sobretudo nos países de línguas germânicas (Alemanha, Holanda, Escandinávia, a parte alemânica da Suíça e a Inglaterra) com a exceção notável da Áustria. Um importante predecessor do movimento protestante foi o tcheco Jan Huss, também queimado na fogueira pelas autoridades católicas em 1415.

Para traduzir a Bíblia, os intelectuais protestantes tiveram de fazer suas opções gramaticais e ortográficas. A importância dessas traduções se reflete no fato de terem servido de base para a constituição da norma-padrão clássica de diversas línguas. A Bíblia de Lutero instaurou a norma do alemão, assim como a tradução de Jurij Dalmatin serviu de base para a norma literária do esloveno. As primeiras traduções completas da Bíblia em línguas europeias foram obras de intelectuais protestantes, como o caso do português João Ferreira de Almeida (1628-1691).

8.3.2 *Idiossincrasias ortográficas*

As línguas diferem muito quanto aos critérios estabelecidos para a definição de suas regras ortográficas. Essas regras são o resultado de um complexo jogo de interesses políticos, religiosos, ideológicos, culturais etc. Uma vez que, como dissemos, a ortografia não faz parte da língua, sendo de fato uma instituição, ela tem estado desde sempre sujeita às arbitrariedades de seus idealizadores.

A função da ortografia é, teoricamente, permitir a transcrição da língua falada. Desse ponto de vista, a ortografia ideal seria aquela que propiciasse uma relação um-a-um entre som e letra. Em nenhuma língua do mundo que utilize a escrita alfabética esse ideal tem condições de ser alcançado, pela simplícissima razão de que ele suporia uma homogeneidade linguística contrária à própria natureza das línguas humanas. Toda e qualquer língua viva apresenta variação, de modo que as opções ortográficas são sempre opções políticas, em que uma variedade linguística — geralmente a da região mais rica do país ou daquela onde está o poder político — serve de base para a fixação das regras da escrita.

No caso das principais línguas românicas, se fôssemos colocar suas ortografias oficiais numa linha imaginária do menos ideal para o mais ideal, teríamos decerto a seguinte situação:

O francês é, de longe, a língua românica que tem a ortografia oficial mais distante do ideal, enquanto o romeno e principalmente o italiano se aproximam em boa medida dele. (O romeno alcançou esse quase ideal porque, em 1860, o alfabeto cirílico foi substituído pelo latino, o que permitiu a criação de uma ortografia nova a partir do zero, mais próxima de uma boa correspondência letra/som.) Uma lista de palavras escritas em francês e em italiano com suas respectivas transcrições fonéticas ilustra essa distância:

◗ francês:	CHANTER [ʃɑ̃'te]	◗ italiano:	CANTARE [kan'tare] ('cantar')
	BEAUCOUP [bo'ku]		MOLTO ['molto] ('muito')
	SAINT [sɛ̃]		SANTO ['santo] ('santo')
	DOIGT ['dua]		DITO ['dito] ('dedo')
	PAUVRE ['povr]		POVERO ['pɔvero] ('pobre')
	RAISON [rɛ'zõ]		RAGIONE [ra'ʤone] ('razão')
	VENGEANCE [vɑ̃'ʒɑ̃s]		VENDETTA [ven'detta] ('vingança')
	COUTEAU [ku'to]		COLTELLO [kol'tɛllo] ('faca')

Como é fácil perceber, as transcrições fonéticas do italiano muitas vezes correspondem à própria forma escrita das palavras, o que é raríssimo acontecer em francês. Uma vez conhecidas as convenções ortográficas do italiano, é muito fácil escrever uma palavra nessa língua valendo-se apenas de sua pronúncia. (É claro que estamos nos referindo aqui ao italiano padrão que, nem de longe, é a verdadeira língua materna dos habitantes da Itália, um país que é um verdadeiro quebra-cabeças linguístico.) Em italiano, ao contrário do francês, as consoantes duplas são pronunciadas como duplas, isto é, longas. A principal dificuldade, para a leitura do italiano, é a identificação do acento tônico da palavra, uma vez que a ortografia estipula a acentuação gráfica somente nas palavras oxítonas: *caffè, università, rococò*. Assim,

palavras como *vivere* ('viver'), *telefonano* ('[eles] telefonam'), *cinema* ('cinema'), *angelo* ('anjo') etc. não trazem acento gráfico.

A ortografia do francês foi estabelecida oficialmente depois da criação, no século XVII, da Academia Francesa, órgão máximo da normatização da língua. Desde então, a escrita do francês passou por pequenas modificações isoladas que, no entanto, mantiveram em grande parte as formas escritas de quatrocentos anos atrás. Esse extremo conservadorismo impediu o acompanhamento das mudanças ocorridas no sistema fonológico da língua:

GRAFIA	PERÍODO	PRONÚNCIA
LOI	séc. X	['lɔɪ]
LOI	séc. XII	['lʊe]
LOI	séc. XVI	['lʊɛ]
LOI	séc. XVIII	['lʊa]

Assim, a palavra escrita LOI ('lei') passou por diversos estágios de mudança em seu ditongo, mudanças que a ortografia oficial não acompanhou. O conservadorismo da ortografia francesa se deve, também, em grande medida, ao fato do francês ser, de todas as línguas românicas, a que mais se distanciou de sua matriz latina. A ocupação do território francês no século V pelo povo germânico chamado franco — do qual o país herdou seu nome — provocou grandes transformações na versão do latim vulgar falado naquele território. Sobre esse latim vulgar a língua dos francos agiu como superestrato [▶134], influenciando profundamente o que mais tarde viria a se chamar "língua francesa". Diversos aspectos fonológicos e morfossintáticos distanciam, desse modo, o francês das demais línguas românicas: (i) a consoante uvular [ʁ], em todas as posições silábicas, o que também ocorre em alemão; (ii) a inversão da ordem SV (sujeito-verbo) na interrogação, inexistente nas demais línguas românicas, mas característica das germânicas, como o inglês (*Are you French?*) e o alemão (*Bist du Franzose?*), aos quais corresponde o francês *Es-tu français?*; (iii) emprego do prefixo *mé-*, *més-*, derivado do frâncico, com sentido negativo ou pejorativo: *méconnaître* ('desconhecer'), *mécontent* ('descontente'), *méchant* ('malvado'), *mésaventure* ('desventura'), *mésalliance* ('aliança indevida'), correspondente ao inglês *mis-* em *misunderstanding* ('mal-entendido'), *misbehavior* ('mal comportamento'), *misfire* ('disparar [arma] de modo inadequado') e, por fim, (iv) a obrigatoriedade que existe em francês da explicitação do sujeito por meio de um pronome quando as demais línguas românicas, pela desinência verbal, permitem a omissão do sujeito — compare o italiano *canto, canti, canta* com o francês *je chante, tu chantes, il chante*: sem o pronome-sujeito seria impossível distinguir a pessoa a que se refere o verbo. Ora, essa é uma característica das línguas germânicas, nas quais também o verbo jamais ocorre desacompanhado de seu sujeito (inglês: *I sing, you sing, he sings, we sing, you sing, they sing*).

Outro aspecto que distingue radicalmente o francês das outras línguas românicas é a acentuação tônica. Nas demais línguas da família, as palavras podem ser oxítonas, paroxítonas, proparoxítonas e até sobresdrúxulas, isto é, com sílaba tônica anterior à antepenúltima, como em italiano *loro telefonano* ('eles telefonam') ou no português clássico *queixávamo-nos*. Em francês, no entanto, todas as palavras são oxítonas, inclusive aquelas com aparência paroxítona: *France* ('França'), embora, para fins de metrificação, seja considerada díssilaba, sua pronúncia corrente, e que vem registrada nos dicionários da língua, é [fʁãs]. Por isso, no momento de pronunciar qualquer palavra estrangeira, os franceses a transformam imediatamente em oxítona: *Ipanemá, Copacabaná, Riô de Janerô, Barack Obamá, Lulá, Los Angelès, Santiagô* etc.

O francês também perdeu a distinção característica das línguas românicas entre palavras terminadas em -o, predominantemente masculinas, e -a, predominantemente femininas. A vogal final da imensa maioria das palavras francesas é -e, em geral não pronunciada ou pronunciada [ə], um "e caduco", como dizem os gramáticos franceses. Por isso, as poucas palavras terminadas em -a são quase todas do gênero masculino, o que é absolutamente estranho para as demais línguas da família. É o caso de alguns nomes de flores como *pétunia, zinnia, hortensia*, de animais como *ara* ('arara'), de países como *Honduras, Guatémala, Nicaragua, Vénézuéla, Nigéria*, todos do gênero masculino em francês. Diga-se de passagem que a vogal mais frequente na escrita e na pronúncia do francês é o E (aberto, fechado ou "caduco"), exatamente como se dá em inglês, ao contrário das outras línguas românicas, que têm no A o seu som mais frequente (a tal ponto que muitos escritores vêm tentando, ao longo do tempo, cumprir a tarefa quase impossível de escrever um texto completo ou mesmo um poema que não tenha a letra A).

Para tentar conservar sua aparência românica e sua filiação ao latim, a ortografia francesa se vincula estreitamente à etimologia das palavras, por mais que a pronúncia delas tenha se afastado radicalmente da forma escrita. A palavra *sept*, por exemplo, mantém seu *p* por causa da etimologia latina (*septem*, 'sete'), embora a pronúncia seja [sɛt]; o mesmo vale para *doigt* ('dedo'), que se pronuncia ['dʊa], mas que tenta espelhar a forma latina *digitu-*; ou para *thym* ('tomilho'), que se pronuncia [tɛ̃], mas tenta reproduzir o étimo latino *thymum*. O francês escrito conserva a vogal dupla Œ principalmente em palavras de origem grega (*œcuménique, œdipien, œdème* etc.), nas quais é pronunciada [e], não havendo, portanto, nada que justifique seu uso. Essa mesma vogal Œ aparece em *œil* ('olho'), palavra que se pronuncia ['œɪ], como em *orgueil* [ɔʁ'gœɪ], sem que nessa última palavra ocorra a grafia Œ.

A ortografia francesa também usa o acento circunflexo sobre determinadas vogais para indicar que naquela sílaba, quatrocentos anos atrás, existiu um S que deixou de ser pronunciado: *même* (< *mesme*, 'mesmo'), *île* (< *isle*, 'ilha'), *coût* (< *coust*, 'custo'), *âne* (< *asne*, 'asno') etc. Como dizem alguns franceses, ironicamente, é um acento *in memoriam*. O francês escrito conserva grafias que nas outras línguas românicas foram

simplificadas, como PH, TH, RH, Y (não semivogal), consoantes duplas que não se distinguem das simples etc. Assim, o que em francês se escreve *chlorophylle* corresponde a *clorofila* em espanhol e português, a *clorofilla* em italiano e a *clorofilă* em romeno.

No entanto, uma simplificação radical da ortografia francesa impediria a distinção entre palavras homófonas, isto é, com a mesma pronúncia, que são frequentíssimas nessa língua devido às drásticas mutações fonéticas nela ocorridas — as palavras *saint, saints, sein, seins, seing, seings, ceins, ceins, ceint, ceint, ceints* têm todas a mesma pronúncia [sɛ̃], embora signifiquem, respectivamente, 'santo', 'santos', 'seio', 'seios', 'assinatura', 'assinaturas', '[eu] cinjo', '[tu] cinges', '[ele] cinge', 'cingido', 'cingidos'... apesar de tudo isso, é possível encontrar na ortografia francesa algumas regularidades que facilitam, senão a escrita, ao menos a pronúncia das palavras: AU e EAU sempre são pronunciadas como [o]; OU, como [u]; AI, como [e] ou [ɛ], segundo a região; OI, como [ua]; o S em final de palavra nunca é pronunciado, a não ser em pouquíssimas exceções etc.

Mas nem mesmo essas poucas regularidades do francês escrito ocorrem na ortografia da língua que tem o sistema de escrita menos ideal e reconhecidamente mais complexo de todo o mundo: o inglês. Isso se deve a uma série de fatores relacionados à própria história da língua inglesa. A grafia atual do inglês (que talvez nem devesse se chamar *orto-grafia*) reproduz em boa medida a pronúncia da língua no século XV. Foi justamente nesse século que a imprensa chegou na Inglaterra, trazida por William Caxton (1415~1422 - ca. 1492). Caxton imprimiu os primeiros livros em inglês, muitos deles traduzidos por ele mesmo. Por outro lado, também foi a partir de meados do século XV que começou a ocorrer uma série de transformações fonéticas no inglês que hoje são conhecidas como a Grande Mutação Vocálica (Great Vowel Shift), que transformou, por exemplo o [i:] no ditongo [aɪ]. Enquanto essas mutações ocorriam, Caxton preferiu imprimir seus livros sem atentar para elas, mantendo-se fiel às pronúncias mais antigas. As decisões do tipógrafo acabaram por se tornar o padrão da escrita do inglês, que permanece praticamente o mesmo até hoje, apesar de meio milênio de transformações. Por isso, ao contrário de outras línguas em que é possível identificar regras ortográficas mais ou menos sistemáticas, em inglês praticamente essas regras não existem, de modo que é preciso aprender ao mesmo tempo a grafia e a pronúncia de cada palavra. Um exemplo famoso é o da combinação de letras OUGH que pode ter nada menos do que dez pronúncias diferentes, entre as quais [ʌf] (*enough*, 'bastante'), [ɒf] (*cough*, 'tossir'), [oʊ] (*though*, 'embora, apesar de'), [u:] (*through*, 'através') etc. Outro exemplo é o da grafia IGH para o que se pronuncia [aɪ], como em *high, sigh, light, right, night, sight* etc. Quando uma palavra começa com duas consoantes que não formam encontro consonantal, a primeira nunca é pronunciada: *know* [noʊ], *knee* [ni:], *psycho* ['saɪkoʊ], *write* ['raɪt], *wrong* [rɔŋ] etc.

Essa característica da ortografia do inglês é que permite a existência, na cultura dos países anglófonos, dos famosos campeonatos de *spelling* ('soletração'). A transpo-

sição recente desse fenômeno cultural para o Brasil revela, além da subserviência à cultura norte-americana (representada também pela patética comemoração do Halloween), uma total falta de percepção do caráter regular da ortografia do PB, em que as possibilidades de equívoco ortográfico são sempre as mesmas e, portanto, altamente previsíveis [▶353].

8.4 A ORTOGRAFIA DO PORTUGUÊS

É costume dividir a história da ortografia do português em três períodos distintos, cada um deles caracterizado por uma tendência no tocante ao modo de grafar:

1. Século XIII a meados do XVI: ortografia fonética
2. Do Renascimento ao início do século XX: ortografia etimológica
3. De 1911 a 2009: ortografia moderna reformada

Antes da normatização do português, que só viria a ocorrer no século XVI, a pequena parcela da população capaz de escrever se valia do alfabeto herdado do latim e tentava com ele dar conta de transcrever os sons da língua galega, que era a língua falada no início desse primeiro período, em que Portugal ainda não tinha começado sua expansão rumo ao sul, que só se completaria em 1249, com a conquista de Faro, sob o reinado de Afonso III, quinto rei português. Nessa tentativa, o principal objetivo era encontrar uma boa correspondência entre som e letra, razão por que essas primeiras iniciativas são chamadas de ortografia fonética. Vamos ver um exemplo:

> Depois dos cento e cinquoenta dias, começarom as auguas de miunguar, e esteve a arca sobre los montes de Armenia. E depois que passarom os quarenta dias, abriu Noe a freesta da arca, e deitou fora o corvo, e foi, e non tornou; e depois deitou a poomba, e non achou hu pousasse seu pee, tornou-se a Noe, e depois de sete dias emviou Noe outra vez a poomba, e ela tornou a hora de Vespera, e tragia hũm ramo verde d'Oliveira; e depois d'outros sete dias emviou Noe a poomba, e ela non tornou mais a ele.
>
> (Bíblia medieval portuguesa, séc. XIII-XIV, Gênesis, cap. 27)

Já nesse período encontramos o uso do dígrafo CH para representar o som [ʧ]. Esse dígrafo, importado da escrita do francês — onde representava o mesmo som na época —, também era empregado no castelhano com o mesmo valor fonético. Para o som [ʃ] o galego (e o castelhano medieval) empregava a letra X. Somente muito mais tarde é que, já em português, a consoante [ʧ] perderá sua oclusão, de modo que CH e X vão convergir na representação de [ʃ]. O galego moderno preservou a diferença, usando CH para [ʧ] e X para [ʃ].

No castelhano antigo, também se usou a letra X para representar [ʃ], como testemunha o nome do país México, que os primeiros conquistadores espanhóis pronunciavam ['meʃiko]. Mais tarde, a consoante [ʃ] desapareceu do sistema fonológico da língua, e o que antes se escrevia com X passou a se escrever com J e a se pronunciar [x], o que se verifica nos pares *deixar/dejar, baixo/bajo, xícara/jícara* que diferenciam o português do espanhol. Em sua gramática de 1492, Antonio de Nebrija ainda grafava *debaxo* para o que hoje se escreve *debajo*.

Em textos de 1250 em diante já aparecem os dígrafos NH e LH, que foram importados da escrita do provençal, língua que influenciou grandemente a literatura primitiva galega e portuguesa no cultivo do trovadorismo. Esses dígrafos não serão adotados pela ortografia oficial galega, que usa até hoje Ñ e LL, por influência do castelhano.

O que mais diferenciava a língua arcaica da atual era a existência de pares de consoantes que se opunham pela presença de uma africada e de uma não africada:

AFRICADA	GRAFIA	NÃO AFRICADA	GRAFIA
[dʒ]	GE, GI	[ʒ]	J
[dz]	Z	[z]	-S-
[tʃ]	CH	[ʃ]	X
[ts]	Ç, CE, CI	[s]	S, -SS-

[Obs.: não se tem a mesma certeza da distribuição complementar de [ʒ] e [dʒ] como se tem dos outros pares, mas é possível supor que o quadro de oposições fosse simétrico como o apresentado acima.]

No período em que as africadas eram vivas na língua, não existia indecisão no momento de escrever, pois a cada grafia correspondia um som diferente, como nesses pares:

paço ['paʦo]	x	passo ['paso]
geral [dʒe'ral]	x	jeito ['ʒeɪto]
cozer [ko'dzer]	x	coser [ko'zer]
tacha ['taʧa]	x	taxa ['taʃa]

Por volta de 1500, a série das africadas já tinha desaparecido.

A perda do traço africado ocasionou o surgimento, no caso das sibilantes, de outra dupla de pares que eram empregados em distribuição complementar (ou seja, onde um era usado o outro não era e vice-versa):

	PRÉ-DORSODENTAIS	ÁPICO-ALVEOLARES
SURDAS	[s] - grafado ç, ce, ci como em *paço*	[ś] - grafado s ou -ss- como em *passo*
SONORAS	[z] - grafado z como em *cozer*	[ź] - grafado -s- como em *coser*

O PB desconhece totalmente as consoantes ápico-alveolares, mas o português europeu, até pelo menos o início do século XVI, tinha essas sibilantes, que são pronunciadas com a ponta da língua (o *ápice*) próxima dos *alvéolos* (região frontal da cavidade bucal logo acima dos dentes). O castelhano padrão atual só conhece a sibilante [ś] que, para ouvidos brasileiros, soa quase como um [ʃ], isto é, um "s chiado", o mesmo ocorrendo no galego moderno. As demais línguas do grupo hispânico apresentam o [s] semelhante ao do PB na palavra *sucessos*. Essas diferenças fonológicas eram acompanhadas pela grafia, de modo que, aqui também, não ocorriam dúvidas no momento de escrever. (No castelhano e no galego, o [ʣ] e [ʦ] acabaram por se transformar na interdental [ɵ], que corresponde ao que em inglês se escreve com TH como em *thanks*, também desconhecida das línguas hispânicas faladas no continente americano, ao passo que [z] e [s] confluíram para [ś].)

Já avançado o século XVI, porém, começam a surgir confusões entre os dois membros de cada par, refletidas na escrita que mesclava Ç com -SS- e vice-versa, Z com -S- e vice-versa (exatamente como acontece hoje com as crianças em fase de alfabetização ou com as pessoas alfabetizadas precariamente). O gramático Pero de Magalhães Gândavo, em 1574, já protestava contra essas confusões:

> As letras que se costumão muitas vezes trocar hũas por outras, em que se cometem mais vicios nesta nossa linguagem são estas que se seguem, convem a saber, *c*, *s*, *z*, e isto nace de não saberem muitos a diferença que ha de hũas às outras na pronunciação. [...] E pera saber como se há de fazer esta diferença, entendão que quando pronunciarem qualquer dição com *c*, hão de fazer força com a língua nos dentes debaixo, de maneira que fique algum tanto a ponta dobrada para dentro, e quando for com *s* porão a língua mais folgadamente pera cima, que fique soando a pronunciação à maneira de assovio de cobra [...]. [P. M. Gândavo, *Regras que ensinam a maneira de escrever a orthographia da língua portuguesa*, 1574]

Como sempre, a campanha dos gramáticos para deter as mudanças linguísticas se revelou inútil, pois naquele mesmo final do século XVI em que era publicada a obra de Gândavo as quatro sibilantes se reduziram a duas, [s] e [z], as mesmas que conhecemos hoje no PB. Essa redução, no entanto, só ocorreu no português europeu do centro-sul de Portugal, região que dita a norma. Nos dialetos do extremo nordeste do país, o sistema de quatro sibilantes permanece até hoje, e é em homenagem a esse pequeno punhado de falantes portugueses que todo o resto do mundo onde se fala uma língua derivada do português europeu do século XVI é obrigado a escrever *caçar* e *cassar*, *coser* e *cozer*, embora não haja nenhuma distinção fonética entre as palavras desses pares. Em outra área, tão pequena quanto a anterior, no extremo noroeste, as quatro sibilantes se reduziram a duas também, mas com preferência pelas ápico-alveolares: é o que em Portugal se chama "s beirão", com referência à região da Beira.

Outra homenagem que o resto do mundo presta aos falantes dos dialetos do norte de Portugal é a manutenção da escrita OU para o que de fato se pronuncia [o] no resto do mundo. Essa monotongação parece datar do século XVII e se generalizou na norma-padrão do português europeu e no VGB, de modo que nessas línguas a escrita OU é, de fato, um dígrafo.

A partir do século XVIII, as vogais finais *-e* e *-o* vão sofrer "redução" (isto é, alçamento) e passam a ser pronunciadas como [i] e [u]. A ortografia, porém, não vai acompanhar essas mudanças, de modo que até hoje escrevemos *ovo* para o que pronunciamos ['ovu] e *ave* para o que pronunciamos ['avi].

O século XVI representa uma guinada importante na história de Portugal e, consequentemente, da língua portuguesa. Nos anos finais do século anterior, tinha começado a expansão marítima dos portugueses com a chegada de Vasco da Gama na Índia (1498) e de Pedro Álvares Cabral no Brasil (1500). Em breve, Portugal vai se tornar uma potência imperial moderna, enriquecendo-se com a riqueza que será espoliada dos povos conquistados e escravizados, como o ouro do Brasil, que tornará o rei D. João V (1689-1750) o monarca mais rico do mundo em sua época, além, é claro, da riqueza constituída pelo comércio altamente lucrativo de escravos africanos, reinventado na modernidade pelos portugueses.

Assim como no resto da Europa, o início da chamada Idade Moderna (simbolizado pela queda de Constantinopla sob os turcos em 1453), vai reacender com vigor um apreço pela antiguidade clássica greco-romana. Esse apreço se mostrará em todos os campos da vida cultural e intelectual: na arquitetura, na pintura e nas artes plásticas em geral, na literatura (com o cultivo dos gêneros literários e formas poéticas da antiguidade) e também nos estudos linguísticos. Com o fim do feudalismo e o fortalecimento dos Estados centralizados, as línguas nacionais serão valorizadas e receberão um forte investimento institucional, com o surgimento das academias de língua, a oficialização dessas línguas como idioma do Estado (com o pioneirismo de Portugal, onde isso já tinha sido feito em 1290) e, em resposta a isso, a gramatização, a ortografização e a normatização das línguas.

Esse apego ao classicismo também foi favorecido pela invenção da imprensa — o primeiro livro impresso em português data de 1489. Começa então o período da ortografia portuguesa conhecido como etimológico, embora o nome não seja de todo apropriado, uma vez que foram muitas também as falsas etimologias que impregnaram o modo de escrever a língua.

Entra em curso o uso dos dígrafos CH (com valor de [k]), PH, RH, TH e do Y nas palavras de origem grega, e de CT, GM, GN, MPT nas palavras de origem latina, sem mencionar, é claro, o uso de consoantes duplas como MM, NN, PP, TT, LL etc. Podemos perceber a grande diferença entre a ortografia usada no século XIII-XV e a usada a partir do Renascimento comparando o texto que vimos acima com o que segue, publicado pelo jornal paulistano *Folha da Noite* no dia 1º de junho de 1921:

> O sr. Antonio da Silveira Salles, moço paulista, residente nesta capital, esteve em nossa redacção, fazendo demonstrações praticas do seu invento — um apparelho auto-instructivo, — que visa a desanalphabetisação de qualquer pessoa, independente de professor. É realmente de grande alcance no momento, o invento que tivemos occasião de conhecer agora que o Governo do Estado se preoccupa, na reforma da Instrucção Publica, com desanalphabetisação

> do maior numero de crianças. É o caso até do sr. dr. Secretario do Interior procurar conhecer e adoptar o methodo em questão desde que julgue util, e praticavel. O novo systema simplifica extraordinariamente a aprendizagem das primeiras letras, elementos de arithmetica, geometria e historia, dependendo, apenas, da sua adaptação. Por meio delle será resolvido o problema do analphabetismo para o qual se voltam, ha tanto tempo, todos os que se interessam pelo futuro do Brasil, que são todos os bons patriotas. Pelo lado economico, o Governo do Estado, adoptando-o, veria immediatamente os seus excelentes resultados, segundo affirma o seu inventor. Um exemplo basta para o demonstrar. Paga-se por um professor do interior do Estado, em lugar remoto a quantia de 300$000 mensaes ou sejam 3:600$000 annuaes. O sr. Silveira Salles affirma que se adquirindo 30 desses apparelhos, poderão ser desanalphabetisadas milhares de pessoas, com a decima parte da quantia despendida até então. E não pareça, diz elle, que o apparelho venha tirar a profissão ou diminuir os lucros do professor. Muito pelo contrario, o auxiliará no que ha de mais difficil, podendo com o seu uso desanalphabetisar em quinze dias, quem quer que seja. O inventor dispõe-se a fazer demonstrações praticas a quem desejar conhecer o apparelho — auto-instructivo, — sendo encontrado diariamente á Rua 15 de Novembro n. 34, sobrado, sala 9 de 12 ás 16 horas.

Esse movimento conservador em Portugal ia na contramão do que acontecia na Espanha, onde a Real Academia, no século XVIII, promovia uma reforma da ortografia com base em critérios fonéticos racionais, enquanto a Academia Real das Ciências de Lisboa instaurava a etimologia como o critério ortográfico mais importante. Esse movimento pode ter sido provocado pela forte influência da cultura francesa sobre a portuguesa ou, mais provavelmente, pelo desejo dos intelectuais portugueses de darem à sua língua a aparência mais diferente possível da espanhola. Portugal, desde sua fundação como reino, sempre experimentou o receio de ser reconquistado pela Espanha, vizinha maior e mais poderosa. Daí o esforço permanente por se diferenciar dela.

Outros exemplos dessa ortografia sufocada por penduricalhos inúteis são:

> architectura • assumpto • athmosphera • bocca • chimica • chlorophylla • creança • cysne • Deos • diphthongo • edade • egreja • epocha • exgottar • exhibir • gymnasio • ichthyologia • martyr • orthographia • phantasma • prompto • psychologia • rhythmo • sahida • sciencia • synthese • triumpho • typographia

A proclamação da república em Portugal (1910) foi acompanhada de uma vontade de, junto com um sistema de governo mais moderno, instituir também uma ortografia mais moderna, prática, simplificada. Com isso se inicia o terceiro período da história ortográfica da língua, o período das reformas. A nova ortografia portuguesa de 1911 fez uma ampla e irrestrita "limpeza" no modo de escrever a língua, eliminando todos os sinais gráficos desnecessários ou francamente errados do ponto de vista etimológico (como o TH de *thesoura* ou o H de *hombro*, grafias que nunca existiram em latim). Essa modernização, porém, não foi seguida pelo Brasil: por não ter havido participação brasileira nas decisões ortográficas de 1911, a velha ortografia etimológica perdurou em nosso país até nada menos do que 1945, quando Brasil e Portugal promoveram um primeiro acordo. Para tornar ainda maior a convergência entre a ortografia brasileira e a lusitana, foram feitas algumas novas pequenas alterações em 1971, como a elimina-

ção de vários acentos diferenciais (*jôgo/jogo, flôres/flores, comêço/começo, êste/este* etc.) e do acento grave para indicar um acento subtônico em palavras derivadas (*cafèzinho, sòzinho, pèzinho, possìvelmente, inùtilmente* etc.) Por fim, a partir de gestões iniciadas em 1986 e incentivadas, entre outros, pelo filólogo Antônio Houaiss (1915-1999), foi adotado no Brasil no dia 1º de janeiro de 2009 o Acordo Ortográfico que prevê uma unificação da ortografia para todos os países de língua oficial portuguesa.

8.5 O novo Acordo Ortográfico e a desnecessária polêmica

A adoção do Acordo Ortográfico suscitou, ao menos no Brasil, uma série de mal-entendidos devidos, como sempre, à falta de informação a respeito dos fenômenos linguísticos, principalmente da parte dos meios de comunicação. O primeiro grande equívoco se expressou na forma de uma suposta "unificação da língua portuguesa".

Existe no senso comum a equiparação entre língua e ortografia, quando, de fato, a ortografia não faz parte da língua. Só essa concepção distorcida é capaz de explicar o equívoco tantas vezes repetido acerca da "unificação da língua portuguesa". O que se buscou unificar com o Acordo foi pura e simplesmente a ortografia, o modo de escrever a língua e não de pronunciá-la, muito menos de empregá-la — ninguém jamais pensou numa "unificação fonética" nem numa "unificação morfossintática" da língua, porque essa tarefa é simplesmente impossível: a heterogeneidade é da natureza de qualquer língua humana, de modo que a variação é inevitável. A única possibilidade de unificação se dá, portanto, no modo de escrever, justamente porque a ortografia... não faz parte da língua.

Outro equívoco recorrente é a denominação de "reforma" para as alterações estipuladas pelo Acordo Ortográfico. Não é possível chamar de reforma um conjunto de modificações que afetam tão somente 0,5% das palavras escritas do PB e 1% das palavras do português europeu. As verdadeiras reformas foram as de 1911 e 1945, que modificaram profundamente a aparência escrita da língua. O Acordo Ortográfico apenas aparou algumas poucas divergências que ainda vigoravam entre as grafias portuguesa e brasileira, como o uso do trema e do acento nos ditongos EI e OI, ambos eliminados.

Adaptar-se às novas regras de ortografia é simples e rápido, sobretudo porque elas promovem eliminações, ou seja, coisas que deixaremos de fazer ao escrever. Muito menos traumático do que, por exemplo, ser obrigado a abandonar todo um sistema de escrita em favor de outro, como aconteceu na Turquia em 1928, quando o alfabeto árabe foi substituído pelo latino: do dia para a noite, toda a população do país se tornou analfabeta e teve de reaprender a escrever.

Infelizmente, como escreveu Carlos Alberto Faraco, "parece que alguns filólogos não conseguem mesmo viver sem cultivar alguma picuinha..."[1] O autor se refere especificamente à irracional proibição, no Acordo Ortográfico, da contração da preposição de com os determinantes (como em *o fato* do *Brasil ser um pais continental*). Mas a mesma avaliação pode ser feita ao tratamento dado ao hífen nas novas regras de ortografia.

Excetuando-se os empregos obrigatórios por razões morfossintáticas (*disseram-me, compre-o, dir-se-ia* etc.), o hífen poderia ficar ao gosto do freguês. Quando abrimos um dicionário de inglês, encontramos as grafias *online* e *on-line*, *up to date* e *up-to-date*, *non-return* e *nonreturn* entre outras: tanto faz escrever com hífen ou sem hífen, juntando as palavras numa só ou deixando separadas... Não podia ser assim também em português?

Qual a justificativa para agora escrever *mula sem cabeça* sem hífen, quando antes era *mula-sem-cabeça*? E por que antes se escrevia *microondas* e agora tem que ser *micro-ondas*? Qual a razão para escrever *afro-brasileiro* e não *afrobrasileiro*? Por que tenho de escrever *protoindo-europeu* e não *proto-indo-europeu* ou, melhor ainda, *protoindoeuropeu*? E, principalmente, por que não deixar que todas essas formas sejam usadas sem medo nem estardalhaço?

As línguas românicas (italiano, francês, espanhol, português) sofrem há séculos com a tradição purista e academicista, que impõe sobre a língua uma paranoia de correção, um monte de regras anacrônicas que impede os cidadãos de se apropriar da língua, de sentir seu idioma materno como algo que lhes pertence de absoluto e pleno direito. No caso de nações surgidas do processo colonial, como o Brasil, esse mal se transforma em doença crônica, já que o ideal de correção está sempre do outro lado do mundo. Já nos países de língua inglesa, impera uma repulsa às academias de língua, que nunca brotaram na Inglaterra nem nos Estados Unidos. O resultado é que por lá, mesmo havendo (como em todo canto) os reacionários da língua, a aceitação das formas inovadoras é muito mais rápida e tranquila: a partir do momento em que essas formas começam a aparecer com frequência na fala e na escrita das pessoas mais letradas, elas automaticamente entram para o rol do que é certo e aceitável.

O emprego do hífen pode ter, por parte de quem escreve, uma função importante de destacar um elemento da palavra, uma função, portanto, pragmático-discursiva: uma coisa é simplesmente *rever* ou *refazer*, outra coisa, bem diferente é re-ver, re-fazer. A eliminação do hífen nas palavras compostas com não impede a diferenciação que a pessoa que escreve pode desejar fazer entre a simples negativa de um ato e a negação de um conceito: uma coisa é não querer, outra coisa é o não-querer; não poder e não-poder são ideias profundamente distintas e, para marcar essa distinção, é imprescindível que não-poder seja graficamente visível como uma palavra só.

---

[1] http://www.parabolaeditorial.com.br/novoacordo2.pdf, acesso em 20 agosto de 2011.

Posso escrever afrobrasileiro se quiser mostrar a impossibilidade de separar o elemento africano na constituição da nossa cultura, mas também posso escrever afro-brasileiro para destacar a participação do elemento africano na nossa formação sociocultural... Posso? Não posso, porque — ao contrário do que tanto se esperava — o novo Acordo, em vez de resolver para sempre os seculares problemas do uso do hífen, só veio substituir os velhos problemas por novos.

Por isso, nesta gramática, em postura assumida de "desobediência civil", optei por escrever não-marcado e não-pessoa, porque somente essas grafias conseguem traduzir minha intenção teórica ao optar por esses conceitos. Há tantas coisas importantes sobre as quais legislar nos países de língua oficial portuguesa — entre os quais alguns dos mais pobres do mundo — que bem se poderia ter deixado em paz esse ínfimo sinal gráfico e, assim, deixar em paz tantos milhões que ainda não sabem ler nem escrever.

Uma verdadeira reforma com vistas a uma real simplificação da ortografia teria de escolher, por exemplo, uma única letra para grafar o som [s], que hoje escrevemos com Ç, CE, CI, S, SC, SÇ, SS, X, XC, XS; para grafar o som [z], que se escreve Z ou -Z-; para grafar o som [ʃ], que escrevemos com CH e X, e assim por diante. Isso exigiria decisões radicais e, sem dúvida alguma, um debate interminável em torno de que símbolo gráfico escolher para representar cada som. Cariocas e lisboetas, por exemplo, poderiam exigir que se escrevesse ixcola, ixtrela, fexta, cuxto, porque é assim que se pronuncia o [s] em sílaba travada no seu dialeto... Grande parte dos brasileiros poderia reivindicar uma escrita que contemplasse a palatalização de [d] e [t] diante de [i]: djiantchi, dentchi, verdji? Os cariocas poderiam escrever djixtchinu? E como dar conta, graficamente, das múltiplas realizações possíveis do que escrevemos com a letra R ou com o dígrafo RR? E o que fazer com os falantes dos dialetos do norte de Portugal que ainda conservam as quatro sibilantes que vimos mais acima? A ideia de uma ortografia totalmente desprovida de ambiguidades e redundâncias é um dos muitos mitos culturais que envolvem as questões linguísticas.

8.6 Mitos em torno da ortografia

O mito de que a "escrita representa a fala" e, *por conseguinte*, "é preciso falar como se escreve" ainda constitui um dos principais entraves para uma pedagogia de língua materna no Brasil.

Mesmo que nossa ortografia esteja muito distante das irracionalidades evidentes da francesa e, sobretudo, da inglesa, uma análise mais minuciosa faz emergir uma série de inconsistências. Vamos examinar, por exemplo, as que dizem respeito ao usos dos diacríticos.

O til tem uma função ambígua, que raramente é explicitada pelas gramáticas normativas e pelos livros didáticos. Ora ele indica apenas uma nasalização — *órfã, ímã, bênção* —, ora indica nasalização mais sílaba tônica — *maçã, afã, hortelã* —, mas somente quando a palavra é oxítona e terminada em -ã (e nunca em [ẽ], [ĩ], [õ] e [ũ]), pois em proparoxítonas se usa o acento circunflexo — *âncora, cântico, gôndola, idêntico,* se a vogal tônica for A ou O, pois se for I e U se usa o acento agudo (*íntimo, último*) — e em paroxítonas, nenhum diacrítico: *campo, ponte, santo,* bastando como indicador da nasalidade o M ou N. Mas se a paroxítona termina em consoante, volta o circunflexo — *câncer, âmbar, cânon, plâncton, ônix* —, desde que essa consoante não seja M ou S — *ontem, antes, ambos* —, embora, havendo um U antes do S seja necessário de novo o circunflexo: *ânus, ônus, bônus.* Além disso, existe o caso único de *cãibra / câimbra,* que admite dupla grafia. O til também é usado sobre a letra O, mas somente no grupo final -ões, sempre em sílaba tônica de palavra oxítona no plural: *corações, leões, limões.* Não é muito claro por que se pode escrever *maçã,* mas não *maçõ* (*maçom*), *cã* mas não *cõ* (*com*), *sã* mas não *sõ* (*som*), grafias, por sinal, existentes no século XIII. Outras divergências ortográficas no uso do til também são difíceis de explicar: por que a dupla grafia *acórdão/acordam* ou *gólfão/golfam,* em que a primeira palavra tem acento agudo e til, se em outros pares eles não ocorrem, como em *amaram/amarão*? Por que não *amárão/amarão*? A diferença do uso do til aqui não se deve nem a "som fechado", "som aberto", "som nasal" ou sílaba tônica, mas ao fato de *acórdão* e *gólfão* serem substantivos e *acordam* e *golfam* serem verbos. Mas no par *o verão/eles verão* temos a mesma oposição substantivo/verbo, sem duplicidade de grafia. As regras do uso do til, portanto, encheriam mais de uma página.

Os acentos agudo e circunflexo acumulam duas funções: indicar vogal aberta e fechada, respectivamente, e também a sílaba tônica, como em *pêssego, légua, café, fôlego, óculos.* No entanto, se o circunflexo é usado para indicar vogal fechada, por que não é usado sobre o I e o U, que são sempre fechadas, em lugar do acento agudo para indicar sílaba tônica: *único, túmulo, íntimo, caí, baú, fígado*? Pela própria definição dos usos oferecida pela tradição de ensino, deveríamos escrever *ûnico, tûmulo, întimo, caî, baû, fîgado*... Em outros casos, o circunflexo tem valor exclusivamente diferencial: *ele tem/eles têm; ele vem/eles vêm* — a pronúncia é a mesmíssima e o acento não indica "vogal fechada". Mas por que usar o circunflexo nesses casos se, em ocorrências do grupo -em acentuado, o que se emprega é o acento agudo: *alguém, ninguém, porém, também*? Por que escrevemos *país,* com acento agudo e S, e *raiz* sem acento e com Z no singular, mas com acento no plural (*raízes*)?

Quanto ao acento grave, ele não deveria mais ser chamado de "acento", uma vez que tem mera função gramatical e não indica sílaba tônica. O ensino de seu emprego é um dos principais problemas enfrentados pelos professores (eles mesmos, com muita frequência, não sabem ao certo quando usar esse sinal). Isso se deve a dois fatos: primeiro, não existe diferença fonética audível entre o A simples e o À "craseado" na pronúncia brasileira; em seguida, o uso da preposição *a* está há bom tempo em franco declínio no VGB, inclusive entre os falantes mais letrados [▸867]. Seria

tudo muito simples e fácil se pudéssemos (como sugeriu, aliás, José de Alencar já no século XIX) usar a grafia à para toda e qualquer ocorrência da preposição *a* — na verdade, seria preciso apenas autorizar legalmente esse uso, já que ele ocorre efetivamente na imensa maioria dos usos escritos da preposição, basta andar pelas ruas ou consultar a internet para ver. Mas enquanto os 190 milhões de brasileiros tiverem de representar na escrita fenômenos fonéticos que só ocorrem no português europeu (10 milhões de falantes), essa legalização será impossível.

Se fosse necessário "falar como se escreve", teríamos de inventar um som para a letra H, não poderíamos nasalizar a sílaba tônica de muito, ninguém poderia palatalizar [t] e [d] antes de [i], toda letra L teria de ter pronúncia lateral e não vocalizada, não poderíamos pronunciar como [z] o S intervocálico de *rosa*, teríamos que pronunciar as vogais finais E e O como [e] e [o], sem "reduzi-las" a [i] e [u] etc. Também é preciso lembrar que nós produzimos variações fonéticas automáticas, inconscientes, que decorrem do ambiente fonético e/ou morfossintático em que os sons se encontram, sem que a grafia acompanhe: o S em *gastar* é um [s], mas o S em *rasgar* é um [z], assim como o Z em *faz calor* é um [s]; o M de *rama* é um [m], mas o M de *rampa* é mero índice de nasalização da vogal anterior ['hãpa]; o E de *selo* poder ser um [e] ou um [ɛ] dependendo da classe gramatical da palavra; o O tônico de *porco* é um [o], mas o de *porcos* é um [ɔ], e é a flexão de número que indica isso...

A ortografia oficial, ao servir de sistema único para representar todas as variedades da língua, exibe uma falsa aparência de "neutralidade" que está longe de ser o que realmente determina sua elaboração e instituição. Mesmo as ortografias mais próximas do "ideal" estão sujeitas aos diversos critérios que se confundem no momento de determinar as formas oficiais de escrever:

- o critério fonético → as letras B, F, P e V representam sempre o mesmo som no PB, havendo portanto uma correspondência um-a-um entre letra e som;
- o critério fonêmico → a letra L representa uma abstração, o fonema /l/, que se realiza concretamente no PB como [l], [ɬ], [ʊ] ou [ɹ], de acordo com sua posição na palavra e/ou com a origem regional do falante;
- o critério morfofonêmico → o morfema -S de plural é um [s] em *meus cães*, mas é um [z] em *meus gatos*, podendo também ser um [ʃ] ou um [ʒ] nas variedades que apresentam o "S chiado";
- o critério etimológico → escrevemos *mau*, com U, porque provém de *malu*, com síncope do [l] intervocálico; mas escrevemos *mal* com L porque provém de *male*, com apócope do [e], morfema latino formador de advérbios — no entanto, as duas palavras, na grande maioria das variedades linguísticas brasileiras, têm pronúncia idêntica; o emprego do H inicial só atende ao critério etimológico; as grafias *sessão*, *seção*, *cessão* também se devem exclusivamente à etimologia, assim como *cesta* e *sexta*;
- o critério anacronístico → apesar das reformas ocorridas ao longo do século XX, ainda escrevemos as vogais finais -E e -O, tal como eram pronunciadas antes do século XVIII; o falso ditongo OU, que também desapareceu da norma-

-padrão portuguesa e de todo o PB no mesmo período, e o CH que nas palavras escritas até o século XV apresentavam a consoante [ʃ] e que hoje poderiam tranquilamente ser escritas com X;

- o critério sociolinguístico → a ortografia oficial se baseia em determinada(s) variedade(s) linguística(s) e não em outras, por isso não existe uma grafia para o "R caipira" [ɹ], para as vogais pretônicas abertas das variedades nordestinas ([tɛlɛˈfõni], *telefone*), para o [ʃ] cuiabano de [ˈʃuva] (*chuva*), para o "S chiado" presente em diversas variedades do PB, para o [ʤ] amplamente generalizado no PB ([ˈʤia]) e assim por diante.
- o critério político-ideológico → desde o surgimento da imprensa na Alemanha, os substantivos em alemão são sempre grafados com letra inicial maiúscula, mesmo que sejam substantivos comuns: *die Frau* ('a mulher'), *das Haus* ('a casa'), *der Weg* ('o caminho'). Essa regra era seguida também no norueguês. No entanto, depois que os exércitos nazistas ocuparam a Noruega por cinco anos, os noruegueses deixaram de usar a inicial maiúscula à moda alemã. Já vimos acima por que croatas e sérvios, falantes da mesma língua, usam alfabetos diferentes.

Esses critérios variam de caso para caso, de palavra para palavra, e também se confundem: continuamos a escrever *caçar* e *cassar* por causa da etimologia de cada verbo (*captiare* e *cassare*, respectivamente), porque a grafia diferente equivalia a uma oposição fonética que desapareceu no século XVI e porque essa oposição, embora minoritária e ausente da norma-padrão do português europeu, ainda se preserva em alguns dialetos do noroeste de Portugal — uma mescla dos critérios etimológico, anacronístico e sociolinguístico.

A "neutralidade" da ortografia, portanto, é mais um mito (assim como a ilusão de que é possível "falar sem sotaque", [▶338]), pois o que seu exame revela é, muitas vezes, uma clara arbitrariedade nas decisões tomadas para oficializá-la, como se viu no caso dos acentos agudo e circunflexo. Nada justifica, por exemplo, que se escreva *extravagante*, *extrovertido*, *extraordinário* com X, mas *estranho* e *estrangeiro* com S, se na etimologia dessas palavras todas está o elemento *extra-* do latim. Nem que se escreva *estender* com S e *extensão* com X, se são palavras da mesma família, assim como *exprimir* e *espremer* que são meras formas divergentes do latim *exprimere*. O mesmo vale para *espraiar*, *escarrar*, *esmurrar*, *escavar*, *esfiapar*, *escusar* que, pelo rigor etimológico, deveriam ser escritas com ex-.

Por isso, no processo de alfabetização, o agente alfabetizador precisa ter plena consciência de que nem tudo o que se escreve se pronuncia, nem tudo o que se pronuncia se escreve, e que a ortografia é um conjunto de símbolos, em boa medida arbitrariamente escolhidos, empregados para escrever, os quais, como todo símbolo, exigem um conhecimento prévio para sua interpretação, uma iniciação, uma vez que seu significado não pode ser deduzido apenas de sua figura. Nem tudo que parece um O soa como [o]. Por isso, a ortografia nunca deve ser considerada como um "retrato fiel" da língua falada.

Também é preciso estar ciente de que não é a escrita que determina a fala, mas exatamente o contrário: para tentar registrar a língua falada é que surgiu a escrita. E não há nada na escrita que faça dela uma entidade supostamente mais importante e mais organizada do que a língua falada. Essa suposta importância é fruto exclusivo de fenômenos socioculturais e político-ideológicos relacionados à natureza grafocêntrica das sociedades ocidentais.

Ao tratar desses temas, a linguista Miriam Lemle (1993: vi-viii) escreveu, sem rodeios:

> Estes problemas muito pouco apetecíveis politicamente e muito pouco glamourosos intelectualmente formam, na verdade, o núcleo do tema da alfabetização. [...] [Eles] mostram bem o quanto é complicada a trama da relação simbólica entre letra e som. Um alfabetizador que não tenha os olhos abertos para as intricações dessa trama vai fatalmente falir. Vai falir por razões tão pequenas quanto a de não ver com clareza o fato de que a maior parte das letras representa um dado som num dado contexto, mas outro som em outro contexto, o fato de que há partes arbitrárias no sistema [...]. Vai falir, enfim, pelo simples fato de não se ter dado conta da estrutura da rede de relações simbólicas que lhe cabe ajudar o aprendiz a dominar.

Muitas opiniões foram emitidas e publicadas logo depois da adoção da nova ortografia no Brasil. A maioria delas, infelizmente, sem nenhuma boa fundamentação em seus argumentos. Algumas alegavam, por exemplo, que a eliminação do trema levaria as pessoas a pronunciar *linguiça, aguentar, quinquênio* etc. como se nessas palavras existisse um [g] ou um [k] e não um [g^{w}] ou um [k^{w}]. Ora, os portugueses aboliram o trema em 1945 e nenhum deles se equivoca na hora de pronunciar tais palavras. Além disso, sempre houve hesitações na pronúncia de algumas dessas palavras, como *líquido, liquidação, liquefazer, questionar, sanguíneo, antiguidade*, além de inconsistências como *equino* e *equestre*, que tinham trema, mas *equitação*, que nunca teve. Com a abolição do trema, essas inconsistências desaparecem e as pessoas podem falar aquelas palavras como bem lhes aprouver. Também se alegou que a supressão do acento agudo de palavras como *ideia, geleia, plateia, estreia, Coreia* etc. ia levar as pessoas a confundir o timbre aberto das sílabas tônicas dessas palavras com o timbre fechado de *correia, passeia, cadeia* etc. No entanto, nunca existiu acento diferencial em *velha* e *ovelha*, nem em *colher* (subs.) e *colher* (verbo), nem em *posso* e *poço*, e ninguém se confunde na hora de pronunciar essas palavas. Depois de 1971, eliminamos os acentos que diferenciavam o substantivo do verbo em inúmeros pares como *rêgo/rego, sossêgo/sossego, carrêgo/carrego, trôco/troco, sôco/soco, jôgo/jogo, fôrça/força, almôço/almoço* etc., eliminação totalmente justificável visto que, aqui também, nenhum falante nativo da língua confunde essas palavras. Vale sempre lembrar: a ortografia do inglês não emprega nenhum sinal diacrítico, mas nenhum falante da língua confunde *do* [du:] e *go* [gou], *head* [hɛd] e *bead* [bid] nem *read* [rid] e *read* [rɛd] etc. O aprendizado da língua falada precede o da língua escrita — se a pessoa um dia chegar a aprender a ler e a escrever. É o conhecimento prévio que os falantes têm de sua língua que vai permitir que eles, no processo de aquisição da leitura e da escrita, identifiquem a palavra *viva e sonora* que está disfarçada sob a máscara da letra morta e muda.

Outra ideia sem fundamento — ou seja, mais um mito — é a de que uma ortografia "simplificada" facilita a aprendizagem da leitura e da escrita. Ora, se fosse assim, as línguas mais faladas e escritas internacionalmente — o francês e, ainda mais, o inglês —, com suas ortografias ilógicas e complicadas, jamais teriam alcançado a difusão que alcançaram. A realidade nos mostra que é preciso inverter esse mito: é a educação de qualidade que leva um povo a se apoderar de seu patrimônio letrado e a ser capaz de ler e de escrever bem, independentemente do tipo de sistema de escrita empregado.

Os Estados Unidos e a Grã-Bretanha apresentam índices de alfabetização de 99% da população, *apesar* da ortografia caótica do inglês. A França exibe o mesmo índice, assim como o Canadá, país que tem no inglês e no francês suas línguas oficiais e de ensino. O Japão tem um índice de alfabetização que beira os 100% — no entanto, escrever em japonês é um aprendizado constante: um cidadão japonês letrado aprende em média um ideograma novo por dia; e a escrita japonesa se vale de nada menos do que três sistemas de escrita diferentes: os ideogramas de origem chinesa (*kanji*) e dois silabários (*katakana* e *hiragana*). A China também se destaca como um país onde 93,3% de seus um bilhão e meio de habitantes são alfabetizados e dominam algo entre 3.000 e 4.000 ideogramas diferentes para poderem ler e escrever a contento.

Por outro lado, diversos países que têm o português como língua oficial e de ensino, isto é, uma língua com uma ortografia mais simples e mais regrada que a das línguas supracitadas, apresentam baixos índices de alfabetização: Angola (67,4%), Guiné-Bissau (44,8%), Moçambique (38,7%). No Timor-Leste, só 50% da população é alfabetizada, sendo que no final do período colonial português, em 1975, o analfabetismo chegava aos 90%, mostrando bem a política colonial portuguesa que, ao contrário de outras potências imperiais, nunca se importou com a educação de seus súditos de além-mar. A Guatemala, país onde o espanhol — que tem um dos sistemas ortográficos mais próximos do "ideal" — é a língua oficial, apenas 69,1% da população é alfabetizada (13 milhões de habitantes). Em contrapartida, Cuba, com população semelhante (11 milhões) ocupa o primeiro lugar na lista de todos os países do mundo em qualidade de educação, e seu modelo de alfabetização popular tem sido adotado por outros países latino-americanos. De acordo com os resultados obtidos nos testes de avaliação de estudantes latino-americanos, conduzidos pela Unesco, Cuba lidera, por ampla margem de diferença, nos resultados obtidos pelas 3as e 4as séries em matemática e compreensão de linguagem. Os índices mais baixos obtidos pelos estudantes cubanos superaram de longe os mais altos de outros países do continente, incluindo o Brasil.

Como se vê, é perfeitamente possível alfabetizar todo um povo, desde que haja investimentos consistentes em educação — se eles existirem, a ortografia pode ser de qualquer tipo.

8.7 Como tratar os erros de ortografia

Uma alta porcentagem do que tradicionalmente se chama de "erro de português" em nossa cultura escolar se constitui, na verdade, de desvios com relação à ortografia oficial. Os erros que detectamos na escrita dos alfabetizandos (adultos e crianças) têm de ser analisados sob duas categorias:

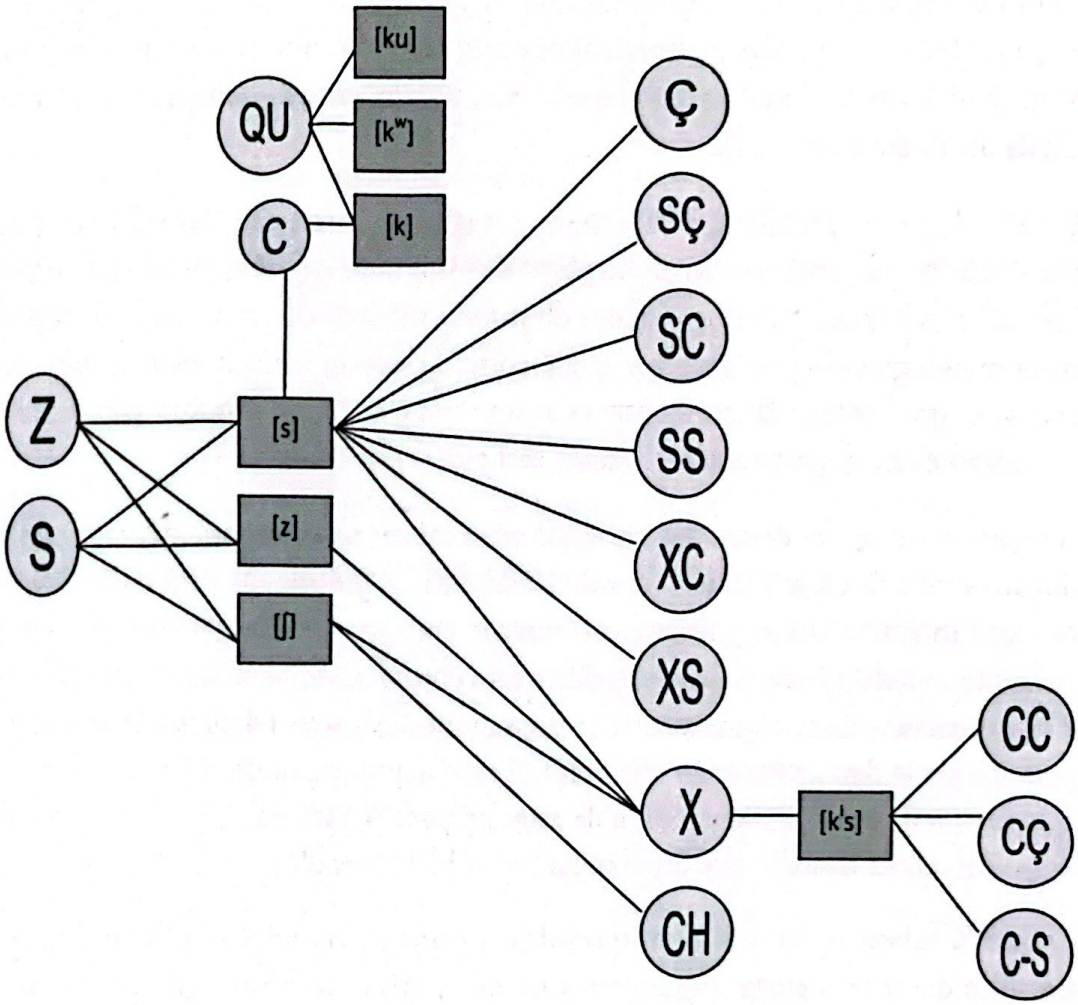
1. hipóteses e analogias sobre a relação fala-escrita que não correspondem às convenções da ortografia oficial;
2. incidência da variedade linguística do indivíduo sobre a produção da escrita.

Ao contrário do que muita gente pensa, a ortografia do português não é estritamente racional e lógica. Ela exibe uma série de incoerências e incongruências que inevitavelmente conduzem os aprendizes a fazerem deduções que, embora lógicas do ponto de vista intuitivo, não encontram respaldo nas convenções (tantas vezes arbitrárias) da ortografia oficial. Retomando algumas dessas convenções, as mais problemáticas (inclusive para pessoas bem letradas), temos esse quadro:

SOM	GRAFIAS POSSÍVEIS	EXEMPLOS
[ʒ]	J	já - fujão - viajem
	G	gente - fugir - viagem
	S*	lesma - rasga - resvala
[ʃ]	CH	cheque - encher - rocha
	X	xeque - enxerga - roxa - texto*
	S*	dois - vistos - meus
	Z*	faz - juiz - arroz
[ʊ]	O	mágoa - coalhada - névoa
	U	mau - cauda - auto
	L	mal - calda - alto
[s]	Ç	aço - poço - iço - ouço - ruço
	C	cinto - cento - sede
	S	sinto - sento - cede
	SC	nascer - crescer - descer
	SÇ	nasça - cresça - desça
	SS	asso - posso - isso - osso - russo
	X	excremento - sexta - trouxe
	XC	exceder - excitar - exceção
	XS	exsudar - exsurgir - exsolver
	Z	faz - juiz - arroz
[z]	S	raso - presa - alisar
	X	exame - exemplo - exímio
	Z	reza - preza - alizar
[kˢs]	X	fixo - tóxico - sexo
	CÇ	facção - ficção - sucção
	CC	faccioso - ficcional - fricção
	C-S	fac-símile

[(*) Nas variedades que apresentam o "s chiado".]

É fácil perceber que os problemas mais numerosos se encontram na multiplicidade de grafias para [s] e nos múltiplos som representados pela letra s. Na verdade, a rede de relações que se estabelece entre os sons [k], [s], [ʃ] e [z] e suas possibilidades de grafia é a mais complexa da nossa língua:



São portanto, 25 correspondências, que podemos exemplificar assim:

1. C [k] — cal
2. C [s] — céu
3. Ç [s] — poça
4. CC [kˢs] — ficcional
5. CÇ [kˢs] — facção
6. CH [ʃ] — chá
7. C-S [kˢs] — fac-símile
8. QU [k] — que
9. QU [ku] — adeque*
10. QU [kʷ] — qual
11. S [ʃ] — testes**
12. S [s] — testes
13. S [z] — uso
14. SÇ [s] — desça
15. SC [s] — descer
16. SS [s] — possa
17. X [ʃ] — luxo, texto**
18. X [kˢs] — táxi
19. X [s] — texto
20. X [z] — êxito
21. XC [s] — exceto
22. XS [s] — exsudar
23. Z [ʃ] — raiz**
24. Z [s] — raiz
25. Z [z] — zoo

[(*) Segundo o dicionário Houaiss, o verbo *adequar* tem duas formas conjugadas: *adéquo* e *adequo*, com sílaba tônica no u.]
[(**) Nas variedades que têm o "s chiado".]

Não é difícil entender as razões das dúvidas, hesitações e incorreções da parte dos aprendizes (e mesmo de pessoas letradas) no momento de escrever alguma palavra que contenha esses sons. Afinal, a aquisição da escrita envolve não só tentar representar a fala por meio de sinais gráficos, mas representar a fala por meio de sinais gráficos segundo as convenções que a tradição ortográfica da língua estabeleceu ao longo dos séculos e que vem sofrendo constantes reformas. A ortografia atual do português, como vimos, reflete, em vários aspectos, a língua falada antes do século XVI...

É natural que o aprendiz faça hipóteses e analogias a respeito das relações entre fala e escrita que, embora lógicas do ponto de vista intuitivo, têm de ser descartadas por não se adequarem às ilogicidades do sistema ortográfico convencional. Não há nada mais lógico do que imaginar que é correto escrever *sinema*, *sebola*, *felis*, *oço*, *essessão*, *xapel*, *otéu*... já que é certo escrever *cesta* e *sexta*, ou *sessão*, *seção* e *cessão*, ou *hesitar* e *exitar*, ou *mau* e *mal*, *meu* e *mel* etc.

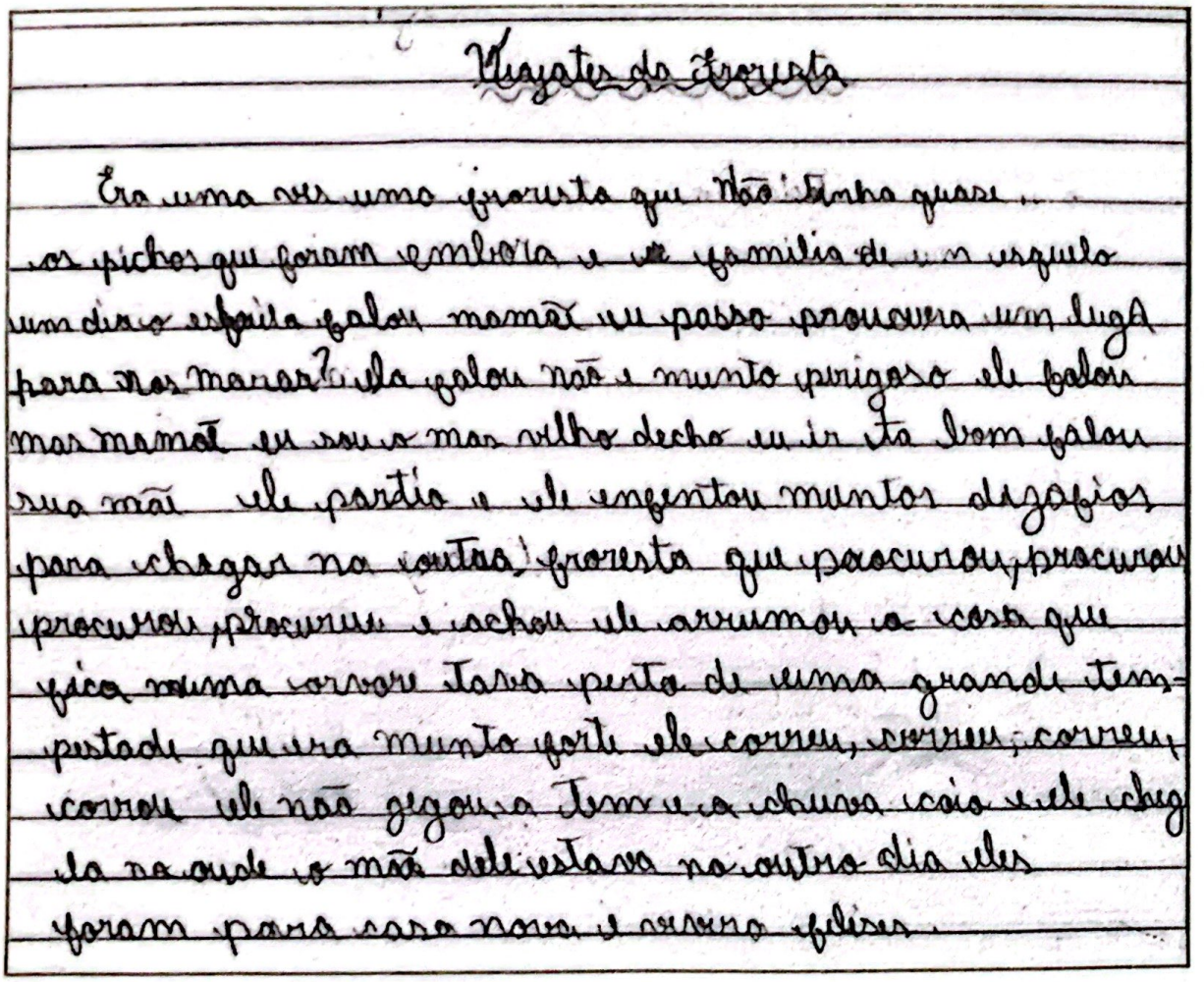
A segunda categoria de desvios ortográficos se relaciona com a incidência, na produção escrita, de características fonomorfossintáticas próprias da variedade linguística que os aprendizes adquiriram no convívio com sua família e na interação com sua comunidade. Como a sociolinguística tem comprovado há bastante tempo, toda e qualquer variedade linguística tem uma gramática, seus falantes seguem regras perfeitamente demonstráveis e coerentes. A escola, privilegiando uma variedade em detrimento de todas as demais, tem de estar preparada para receber falantes de dialetos diferentes daquele que predomina no ambiente escolar.

O que já sabemos há muito tempo sobre a língua e o ensino da ortografia nos permite dizer sem susto que os erros de ortografia são altamente previsíveis. E não poderia ser de outra maneira, uma vez que nosso cérebro é tão organizado e nosso processamento da linguagem é tão eficiente que até mesmo os "erros" são regulares, estruturados, bem organizados... Nenhum de nós jamais ouviu um brasileiro dizer **o peixes morderam a iscas*, mas já ouvimos milhões dizerem coisas como *os peixe mordeu as isca* — afinal, já é uma regra gramatical da nossa língua eliminar as marcas de plural redundantes, desde que a primeira delas tenha sido enunciada para indicar a pluralização de todo o sintagma. Ora, são justamente esses erros que nos mostram o que é e como é, realmente, o funcionamento da língua quando não sujeita ao policiamento consciente das normas prescritivas.

No processo de alfabetização — e aqui pensamos na escola pública e em seu alunado típico, as chamadas "classes populares" —, é imprescindível que a professora conheça os erros mais recorrentes na produção escrita, erros que, vale a pena repetir, são altamente previsíveis. Eles podem ser organizados na seguinte tabela:

ERRO ORTOGRÁFICO	EXEMPLO
Rotacismo: troca do L por R em encontros consonantais, em sílabas travadas ou em final de palavra	bloco → *broco* filme → *firme* falta → *farta* papel → *paper*
Lambdacismo: troca de R por L em sílabas travadas	carvão → *calvão* garfo → *galfo*
Síncope de R em sílaba travada por [s]	comércio → *comecio* diversidade → *divessidade*
Paragoge: escrita de E ou I depois de L final de palavra	sol → *soli, sole* papel → *papeli*
Eliminação das marcas de plural redundante (concordância verbal e nominal)	eles vieram sozinhos → *eles veio sozinho*
Apócope do R em final de palavras (sobretudo verbos no infinitivo, mas também em alguns nomes)	falar → *falá, fala* *professor* → *professô, professo* maior → *maió, maio*
Deslateralização da consoante [ʎ], donde a escrita I em vez de LH	trabalho → *trabaio*
Troca de LH por LI ou vice-versa, por fenômeno semelhante ao anterior; raramente troca de LH por L simples	palha → *palia* família → *familha* mulher → *mulé*
Escrita simples O para o pseudoditongo OU	ouro → *oro* cantou → *canto, cantô*
Escrita simples E do ditongo EI, em determinados ambientes fonéticos	primeiro → *primero, primêro* queijo → *quejo*
Escrita I para o E postônico ou pretônico	menina → *minina* pequeno → *piqueno* pode → *podi* cabe → *cabi*
Escrita U para o O postônico ou pretônico	moleque → *muleque* bonita → *bunita* novo → *novu* falo → *falu*
Síncope da vogal postônica dos proparoxítonos	pássaro → *passo* ônibus → *onbus* sábado → *sabo*
Desnasalização das vogais nasais postônicas.	vantagem → *vantage, vantaje, vantaji* ontem → *onti* eles dormem → *eles dormi, dorme*
Assimilação -NDO > - NO, nos gerúndios	falando → *falano, falanu* dormindo → *durminu*
Escrita U para L	mel → *meu* volta → *vouta* anzol → *anzou*
Epêntese de I antes de S	mas → *mais* capaz → *capais, capaiz*
Escrita do I pronunciado em sílaba travada por consoante	pneu → *pineu, peneu* advogado → *adivogado, adevogado*

Vamos examinar de que forma essas duas categorias de erro ortográfico ocorrem na produção escrita de um aprendiz. O exemplo abaixo é um texto escrito por uma pessoa adulta em processo de alfabetização. Abaixo da reprodução do manuscrito colocamos sua transcrição para facilitar a análise:


Viajates da froresta

Era uma ves uma froresta que não tinha quase os pichos que foram embora e a familia de um esquilo um dia o esquilo falou mamãe eu posso proucura um luga para nos morar?. ela falou não e muinto perigoso ele falou mas mamãe eu sou o mas velho decho eu ir tá bem falou sua mãe ele partio e ele enfentou muntos dizafios para chegar na outra froresta que procurou, procurou procurou, procurou e achou ele arrumou a casa que fica numa arvore tava perto de uma grande tem= pestade que era munto forte ele correu, correu, correu, correu ele não gegou a tem e a chuva caio e ele chg la na onde a mãe dele estava no outro dia eles foram para casa nova e vivero felises.

VIAJATES DA FRORESTA

Era uma ves uma froresta que não tinha quase
os pichos que foram embora e a familia de um esquilo
um dia o esquilo falou mamãe eu posso proucura um luga
para nos morar?. ela falou não e muinto perigoso ele falou
mas mamãe eu sou o mas velho decho eu ir tá bem falou
sua mãe ele partio e ele enfentou muntos dizafios
para chegar na outra froresta que procurou, procurou,
procurou, procurou e achou ele arrumou a casa que
fica numa arvore tava perto de uma grande tem-
pestade que era munto forte ele correu, correu, correu,
correu ele não gegou a tem[po?] e a chuva caio e ele cheg[?]
la na onde a mãe dele estava no outro dia eles
foram para casa nova e vivero felises.

Vamos acompanhar o texto, indicando com o número 1 os fenômenos relativos a hipóteses de grafia com base na experiência já vivenciada de escrita, e com o número 2 os fenômenos decorrentes da influência da variedade linguística do falante sobre sua produção textual:

- viajates[1] — é muito comum que crianças e adultos em processo de alfabetização não escrevam as consoantes M e N quando elas são meros indicadores do traço nasal da vogal anterior. Em seus primeiros contatos com a escrita, o aprendiz supõe (com razão) que a cada sinal escrito deve corresponder um som. Só mais adiante vai descobrir e assimilar o fato de que as vogais nasais podem ser grafadas de diversas maneiras (ver acima os vários dígrafos e trígrafos convencionados para a grafia das vogais nasais). Observe-se, porém, que é só nessa palavra que falta a consoante nasal, pois ela aparece em outros pontos do texto: *enf[r]entou*, *grande*, *tempestade*, *onde* e, mais particularmente, *muinto*.
- froresta[2] — esse é um indício de que o autor desse texto provém de uma comunidade em que os encontros consonantais de oclusiva + /l/ não fazem parte da fonologia da variedade linguística ali falada. Conforme já estudamos, o fenômeno do rotacismo esteve presente na formação histórica da língua, e os falantes de VLE que se valem dele estão simplesmente levando adiante uma tendência muito antiga na história do idioma. Aqui vai ser preciso um trabalho específico da parte da professora para introduzir o aluno não só nas convenções da escrita como também no universo das VUP, nas quais se inspiram as regras ortográficas e gramaticais não estigmatizadas. Nesse caso, portanto, não se trata de erro de ortografia, mas de uma variedade linguística que apresenta outro sistema fonético, diferente do das variedades urbanas de prestígio.
- ves[1] — nossa ortografia apresenta os pares *raiz* e *país*, *francês* e *talvez*, *vês* e *vez*. Nada mais natural do que a pessoa nessa fase ainda não reconhecer os critérios usados pelos instituidores das convenções ortográficas, critérios que podem ser de ordem lexical (adjetivos pátrios sempre terminam em *-ês/-esa*, por exemplo, embora, até 1945, terminassem em *-ez/-eza*), mas também, como vimos acima, etimológicos ou anacronísticos. *País*, por exemplo (que já se escreveu com Z até 1945), se escreve com S por razões etimológicas: sua origem é o francês *pays*.
- pichos[1] — é comum, no início da aprendizagem da escrita, ocorrer troca de letras que representam sons que têm traços em comum, como é o caso de [b] e [p], ambas bilabiais e oclusivas, diferentes somente pelo traço de vozeamento (o [b] é sonoro e o [p] é surdo). O mesmo se dá mais adiante com a grafia *gegou* no lugar de *chegou*: são duas palatais que só diferem, mais uma vez, pelo traço de vozeamento ([ʒ] sonoro, [ʃ] surdo).
- proucura[1,2] — é importante observar que o produtor desse texto não usa nenhum diacrítico, a não ser o til nas palavras *não* e *(ma)mãe*. Também não usa nenhum sinal de pontuação, a não ser um ponto de interrogação uma única vez. Também parece que usou ponto final, mas o texto escrito a lápis não permite uma boa identificação. A intenção aqui era escrever *procurar*. Observe o uso do dígrafo OU, revelando que o autor já tomou consciência de que a vogal [o] pode ser escrita de vários modos. A hipótese é clara: se existe *pouco*, *roupa*, *louco* etc.,

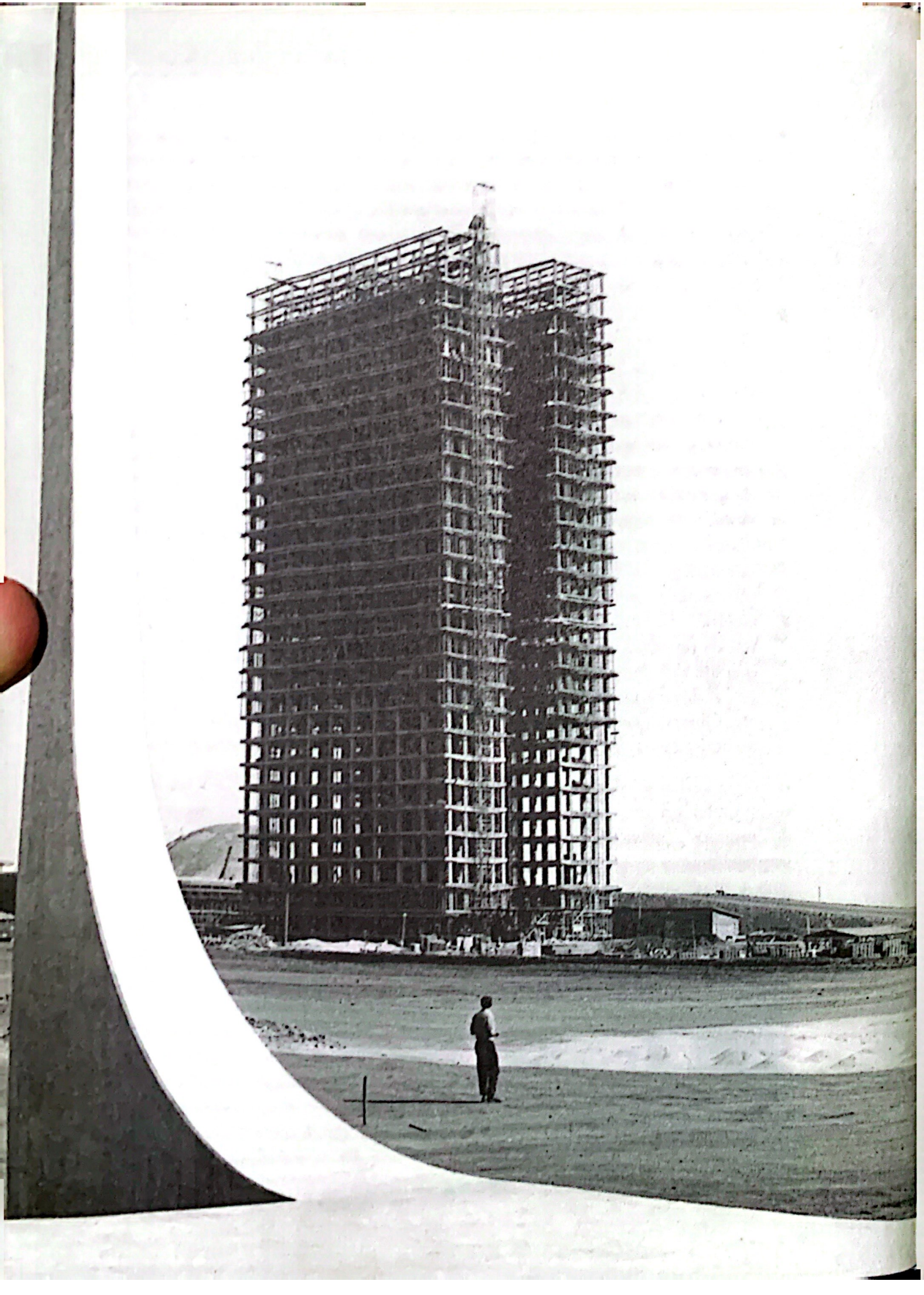
em que o que se escreve OU tem pronúncia [o], é plausível supor *proucurar*. A pronúncia de OU como [o] é generalizada em todo o PB, não sendo, portanto, um traço de variedade linguística específica. Por sua vez, a ausência da letra R demonstra outro fato generalizado no PB, inclusive em suas VUP: a apócope do [h] ou [x] final dos infinitivos e, em geral, das palavras terminadas com essa consoante. Aprender a escrever, portanto, é se dar conta de que a ortografia oficial nos impõe grafar letras que não são pronunciadas, assim como letras que fariam supor determinado som [ou], mas que de fato têm outra realização fonética. A mesma análise vale para a palavra escrita *luga*, que revela a intenção de escrever *lugar*.

- muinto[1] — por influência da consoante [m] inicial, o ditongo [ui] da palavra *muito* se nasalizou em [ũĩ], sendo essa (junto com suas formas no feminino e no plural) uma das duas únicas palavras do PB que contêm esse ditongo (a outra é *ruim* [hũĩ], junto com seu derivado *ruindade*). De novo temos aqui a intuição de que a cada som deve corresponder uma letra, razão por que o autor do texto escreve o N.
- munto(s)[2] — embora na primeira ocorrência da palavra apareça o ditongo, nas outras duas o I não é escrito. Certamente, aqui, a pessoa se deixou levar pelas características de sua variedade linguística, na qual, como em muitas outras do PB, ocorreu o apagamento da semivogal. Vale lembrar que em algumas variedades linguísticas a palavra *muito* é pronunciada ['mũʧu], com palatalização do [t] que se segue ao [i]. Palavras que no português arcaico apresentavam o ditongo UI deixaram de tê-lo em fase posterior da língua: *fructu- > fruito > fruto; lucta > luita > luta; conductu- > conduito > conduto*. É provável que o desaparecimento do [i] tenha se devido a um processo de *dissimilação total* [▶152], isto é, eliminação de um som de pronúncia muito próxima do outro ([i] e [u] são as vogais mais altas e fechadas que temos). Curiosamente, no galego moderno, essas palavras se tornaram *moito, froita, loita*, também resultado de uma *dissimilação* (parcial dessa vez), em que o [u] se abriu em [o] para tornar o ditongo mais perceptível.
- mas[1] — onde se esperaria *mais* o que se lê é *mas*. A hesitação no momento de grafar a conjunção *mas* é frequente, inclusive entre falantes bem letrados. Como na grande maioria das variedades do PB é comum a inserção de uma semivogal [i] depois da vogal tônica final quando precedida de [s], se tornam idênticas as pronúncias de *paz/pais, pôs/pois, mas/mais*, gerando a hesitação mencionada. A conjunção *mas* é resultante de uma gramaticalização do advérbio *mais*, o que ainda resulta em oscilação semântica da parte dos falantes de PB.
- decho[1,2] — já vimos que o ditongo escrito EI se monotonga diante de consoantes palatais ([ʒ] e [ʃ]) e da vibrante [ɾ]. Também tem pessoas que pronunciam *colheita* sem o ditongo: [ko'ʎeta]. Isso leva o aprendiz a não escrever o I do pseudoditongo. Para representar a palatal [ʃ], optou pelo dígrafo CH, como poderia ter optado pelo X. A hipótese provável é: se existe *fechar*, deve existir também *dechar*. Interessante é a presença do O no final da palavra, indicando a contração, absolutamente normal no PB, inclusive entre os falantes mais letrados, *deixa eu ir > deix'eu ir > deix'o ir*. O uso dos pronomes retos em construções causativas/sensitivas [▶991] é regra morfossintática há muito tempo implantada na gramática do PB. A construção *deixa eu ver*, como sabemos, se contraiu em *xovê* na fala de todos os brasileiros, de todas as regiões e classe sociais. O interessante é que o autor do texto escreveu *decho eu*, pois não tem conhecimento de que em *decho* já está contraído o índice pessoal *eu*.

- partio, caio[1] — a pessoa que escreveu esse texto já percebeu que a vogal escrita O pode com muitíssima frequência ser pronunciada [u], como no final de praticamente todas as palavras da língua do gênero masculino (*gato, garoto, bonito, ovo, livro, brasileiro* etc.) e na 1ª pessoa do singular dos verbos no presente do indicativo (*canto, falo, vejo, minto, ando, vendo, durmo* etc.) entre outras tantas. Talvez até por analogia com essas formas do presente ela possa ter decidido empregar o O no lugar do U (*partiu, caiu*).
- dizafios[1,2] — o alçamento das vogais pretônicas é fenômeno generalizado no PB, sendo seu tipo e sua frequência altamente variáveis de região para região e mesmo de falante para falante. Muitas palavras iniciadas com DE- ou DES- são pronunciadas com [di-]/[ʤi-] ou [dis-]/[ʤis-]: *devagar, depressa, demais, depois, delícia* etc.; *desculpa, destilado, destruir, desinformar, desonrar, desonesto, desconfiar* etc. Temos também a dupla de prefixos des-/dis-, com pronúncia que tende a ser idêntica na fala espontânea e mesmo na fala monitorada de pessoas letradas. Isso explica a dificuldade que muitas apresentam no momento de escrever *descrição, discrição, disenteria, destilar* etc. A grafia *dizafio* se explica dessa forma. Quanto ao Z, é mais uma daquelas letras que entram na malha de correspondências que vimos acima [▶391]. Por analogia com *vazio, fazia, dizia* surge a aposta em *dizafio*. Inversamente, e pela mesma razão, a última palavra do texto, *felises*, aparece com S no lugar de Z.
- vivero[2] — ocorre no PB, em todas as variedades linguísticas, incluindo as urbanas de prestígio, uma tendência à redução do ditongo nasal escrito -ão ou -am quando em sílaba átona — *órfão* ['ɔhfu] —, principalmente na não-pessoa do plural dos verbos no pretérito perfeito — *fizer[u], cantar[u], comer[u], dormir[u]* etc. Observe que nosso autor usou O para representar O [u] átono final, revelando consciência de que não existe palavra paroxítona em PB terminada em U.

Pelo que acabamos de ver, os chamados erros de ortografia são altamente previsíveis e perfeitamente explicáveis. Nosso cérebro é tão organizado que mesmo quando "erramos", "erramos" com uma lógica claramente perceptível, ditada por nossas impressionantes e poderosas faculdades cognitivas.

Como essa rápida análise deve ter deixado claro, é fundamental (para não dizer indispensável) que o alfabetizador, para bem desempenhar a sua tarefa, tenha se apoderado dos postulados básicos da sociolinguística e de um sólido conhecimento da história e dos desdobramentos políticos que levaram à instituição da ortografia da língua da forma como existe hoje e não de outra. Alfabetizar é trabalhar com a linguagem, e não é possível trabalhar com a linguagem sem recorrer às ciências da linguagem. É deplorável que nos nossos cursos superiores de pedagogia, que formam os futuros alfabetizadores, só se recorra a teorias psicológicas e pseudocognitivistas sobre a aquisição da escrita e não se propicie uma formação — repito, indispensável — em linguística, sociolinguística, história da língua, história da escrita etc.



LIVRO I LEXICOGRAMÁTICA DO PORTUGUÊS BRASILEIRO

a Rosa Virgínia Mattos e Silva †

9

um presente de grego

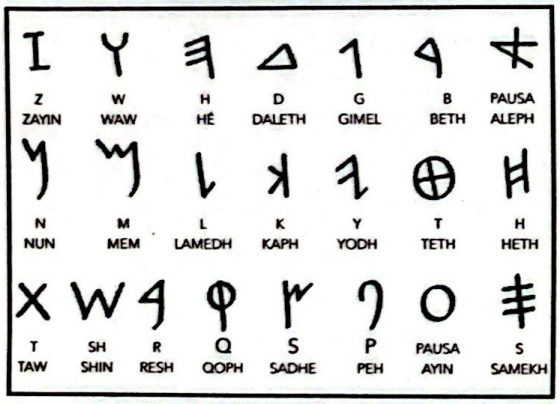
história das classes gramaticais

Como praticamente todas as demais áreas de conhecimento que se instituíram no Ocidente, os estudos linguísticos também são uma herança que nos foi legada pelos gregos.

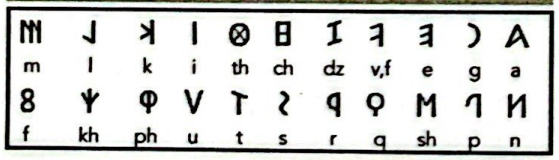
As origens da civilização grega remontam ao século XII a.C., mas esse período inicial é pouco conhecido. A partir do século IX a.C., os gregos adotam e modificam o sistema de escrita inventado pelos fenícios e criam o primeiro alfabeto, que servirá

de modelo para os diversos alfabetos posteriores, com especial destaque para o alfabeto latino que é, hoje em dia, o mais empregado no mundo todo e se originou de uma adaptação do alfabeto etrusco, inspirado no grego. Dizemos que o alfabeto grego foi o primeiro a merecer esse nome porque, ao contrário da escrita fenícia, que só representava as consoantes (tradição que se conserva até hoje nas escritas do árabe e do hebraico, línguas da mesma família do fenício), o sistema de escrita dos gregos incluiu as vogais, o que permitiu a percepção do acento e o tratamento da sílaba (palavra de origem grega) como unidade indivisível das palavras: como muitas sílabas (em grego e em muitas outras línguas) podem ser constituídas de uma única vogal, a notação dos sons vocálicos tornou mais clara a estrutura silábica. O alfabeto grego também permitiu a visualização gráfica das flexões, dando origem aos paradigmas nominais (declinações) e verbais (conjugações) e possibilitando a análise morfológica das palavras, com a identificação das raízes, dos prefixos, dos sufixos, dos morfemas da língua, enfim. Não admira, portanto, que a gramática tenha nascido, no Ocidente, em terras de cultura grega.

ESCRITA FENÍCIA



ALFABETO ETRUSCO



ALFABETO GREGO

A α	alpha		N ν	nu	
B β	beta		Ξ ξ	ksi	
Γ γ	gamma		O o	omicron	
Δ δ	delta		Π π	pi	
E ε	epsilon		P ρ	rho	
Z ζ	zeta		Σ σς	sigma	
H η	eta		T τ	tau	
Θ θ	theta		Y υ	upsilon	
I ι	iota		Φ φ	phi	
K κ	kappa		X χ	chi	
Λ λ	lambda		Ψ ψ	psi	
Mμ	mu		Ω ω	omega	

Entre os séculos VI e IV a.C., a civilização grega atinge uma fase conhecida como Idade de Ouro ou Época Clássica de sua história. É nesse período que vamos encontrar em atividade os grandes nomes da ciência, da filosofia e da literatura gregas que até hoje exercem impacto sobre a cultura ocidental: Heráclito, Protágoras, os sofistas, Pitágoras, Sócrates, Platão, Aristóteles, na filosofia; Sófocles, Ésquilo, Eurípides e Aristófanes, na dramaturgia; Heródoto, Tucídides e Xenofonte, na historiografia; Sólon e Péricles, na política; Hipócrates, na medicina; Anacreonte, Safo e Píndaro, na poesia lírica; Demóstenes e Isócrates, na retórica...

E é também nesse período de apogeu da civilização grega que surgem as primeiras reflexões em torno da linguagem humana em geral e da língua grega em particular

(que, para os gregos, pouco interessados nas demais línguas, às quais chamavam de "bárbaras", eram na verdade uma coisa só).

9.1 A PALAVRA 'GRAMÁTICA'

São os gregos também que vão nos legar a palavra **γραμματική** (*grammatiké*), derivada de *grámma*, termo geralmente traduzido por "letra". É importantíssimo ressaltar, porém, que os gregos (e, depois deles, os romanos) tinham concepções bem particulares da natureza do *grámma* (em latim: *littera*), tido como a menor unidade da fala (latim: *vox*; grego: *phoné*). Existiam duas perspectivas distintas. De acordo com uma, o *grámma* era o símbolo escrito, a representação do som da fala, chamado de "elemento" (grego: *stoikheîon*; latim: *elementum*). Essa posição, precursora da moderna dicotomia letra/som ou grafema/fonema, foi menos importante na Antiguidade (e, de fato, até por volta de 1800) do que a segunda visão, mais complexa. Os filósofos descreviam o *grámma* como uma entidade com três propriedades: nome, forma (ou figura) e som (ou valor). Essa teoria mais flexível, suscetível de extensão e refinamento num grau muito maior do que a mera oposição entre letra e som, foi a base para uma série de abordagens multifacetadas e infinitamente variadas do *grámma* e da *littera* por parte dos filósofos antigos e, mais ainda, dos medievais, que chegaram a ver no estudo da relação letra/som um caminho de acesso à realidade transcendental, mística.

Por isso, embora seja comum dizer que a palavra *grammatiké* (em geral qualificando *tekhné*, "arte", em tekhné grammatiké, "arte da gramática") possa ser traduzida como "estudo da língua escrita", essa tradução só é válida para a gramática alexandrina, surgida no século III a.C. Os filósofos anteriores a esse período — Platão, Aristóteles, os estoicos — usavam o termo *grámma* para designar ao mesmo tempo unidade da fala e unidade da escrita, e elaboraram refinadas teorias filosóficas para explicar a relação entre essas duas realidades, relação que para eles estava longe de ser simples e direta, mas comportava implicações psíquicas, lógicas e até metafísicas.

No atual estágio dos estudos linguísticos, é fundamental distinguir os sons da fala de suas representações na escrita, as letras. Um estudante de linguística, assim que começa seu curso, é (ou deveria ser) instruído a fazer claramente essa distinção e a não se deixar levar pela tradicional confusão entre as duas entidades. E o mesmo vale, é claro, para quem vai ensinar qualquer língua e principalmente a língua materna.

Infelizmente essa confusão vigora até mesmo nos livros didáticos destinados ao ensino da língua. Numa coleção, por exemplo, ao tentar abordar a variação linguística, os autores apresentam um texto escrito numa variedade rural artificial e propõem o seguinte exercício: